# ENIGMA Cerebellum Volumetrics Pipeline Tutorial and Installation Guideline

The initial protocols have been developed by the **ENIGMA-Ataxia working group** (ACAPULCO segmentation pipeline, and SUIT VBM pipeline). Our detailed guideline is based on these previous developments - thanks! This detailed tutorial and guideline has been developed by, Tuğçe Yıldız, Karolinska Institutet, tugce.yildiz@ki.se; and Kristoffer Månsson, Karolinska Institutet, kristoffer.mansson@ki.se.

---

These methods are intended for use in normative samples and clinical populations across the lifespan (except young children), including those with cerebellar atrophy. It is anticipated that they will work with most standard T1-weighted images with full cerebellar coverage. Images should, however, be inspected to exclude clear and obvious artefacts or lesions in the cerebellum.

The pipeline consists of two largely automated modules:

1) ACAPULCO for segmentation of the cerebellum into 28 subunits using a deep learning algorithm, and the use of Freesurfer (autorecon 1) for the estimation of intracranial volume (ICV).
2) SUIT toolbox (with SPM12) for generation of voxel-based morphometry (VBM) maps of the cerebellar grey matter. This module relies on the outputs of ACAPULCO.

In most cases, ENIGMA projects will use both modules, unless otherwise indicated by your project PI. ICV does not need to be re-estimated if already available from other Freesurfer analyses.

**The following detailed instructions will guide you through the analysis and quality control steps for running ACAPULCO, Freesurfer's autorecon1 (for ICV) and SUIT. These modules are run independently of each other (although SUIT uses ACAPULCO outputs).**

**Click here to access the files and scripts necessary for running the pipeline.**

*This guideline is for Linux users. In order to use Acapulco on Linux, you need to install Singularity (Docker is an option).*

**Download and Install Singularity:**

- Download Singularity from: https://www.sylabs.io/docs/
- You can follow the installation steps in the PDF file "SingularityCE User Guide" at this link.

If you face errors starting Singularity, try these:

```
sudo apt-get install -y \
wget \
build-essential \
libseccomp-dev \
libglib2.0-dev \
pkg-config \
squashfs-tools \
cryptsetup \
runc
```

```
export VERSION=3.10.2 &&\
wget
https://github.com/sylabs/singularity/releases/download/v${VERSION}/singularity-ce-${VERSION}.tar.gz && \
tar -xzf singularity-ce-${VERSION}.tar.gz && \
cd singularity-ce-${VERSION}

./mconfig && \
make -C builddir && \
sudo make -C builddir install
```

**Start the pipeline**

*For the pipeline, you should write the following commands on the Ubuntu terminal.*

- In file explorer create a new folder with the name “enigma” and subfolders “acapulco” and “input” in it.
- ACAPULCO Pipeline Files. Download and save the following into your ‘acapulco’ directory: https://drive.google.com/drive/folders/1Wr-7tb0c4vZBGz_cQo1bGNQYW3XiqDsE

    a) acapulco-0.2.1.tar.gz (for DOCKER) OR acapulco-0.2.1.sif (for Singularity)
        - The ENIGMA QC pipeline relies on outputs from version 0.2.1. Please do not use other versions.

    b) QC_scripts.zip: place the contents of this archive into the “acapulco” directory.

    c) R.sif (for Singularity), *OR* calculate_icv.tar (for Docker)

**Directory Structure and Data Set-up**
Create a directory or folder called “enigma”, with “input” and “acapulco” subdirectories. Inside the acapulco directory, create a subdirectory called “output”. The output directory is where the acapulco outputs for each subject will go (see below).

Your ‘input’ folder should contain the T1-weighted scans of all your participants in a nii.gz-format (e.g., subj1.nii.gz, subj2.nii.gz, etc). You can use any prefix you’d like, as long

as they are subject-specific. The below tutorial code assumes this structure, but some deviations won't cause problems (i.e., data in BIDs format, with image files located deeper in the tree) if you tweak the code accordingly.

Place all script files in the acapulco directory.

## ACAPULCO

**Step 1: Running ACAPULCO on your images with Singularity (~5 mins per subject)**
You are now ready to start running acapulco to parcellate the cerebellum! We recommend trying out the pipeline on a single image first. You may need to type "sudo" before the first word "docker" and then provide the system password.

1. Open a terminal/command line window, and change directory to your "acapulco" folder:

```
cd <path/to/acapulco_dir>
```

2. Manually on Windows: Copy your file from the input folder to a subject-specific output folder inside 'acapulco'. The name of the folder should match the name of the image (e.g., folder 'subj01' for file 'subj01.nii.gz'). Write these on Ubuntu terminal:

```
mkdir output/<subject>
cp../input/<subject>.nii.gz output/<subject>/
```

3. Providing singularity is installed correctly on your machine, and you have the acapulco-0.2.1.sif file saved in your acapulco directory, you can run the Singularity pipeline:

```
singularity run --cleanenv -B $PWD:$PWD acapulco-0.2.1.sif –
output/<subject>/<subject>.nii.gz -o output/<subject>
```

Please be aware that the code above is one line code.

Note: Absolute paths to the subject folder for "-i" (input T1 image) and "-o" (output subject folder), will be required if you are not running acapulco from within the 'acapulco' directory. This is what you should see as it's running:

```
[rebeccak@m3p000 acapulco]$ singularity run acapulco-0.2.1.sif -i con006/con006_t1.nii.gz -o con006/
N4 correction...
MNI registraion...
 bad det -1 v 1 u -1
 bad det -1 v 1 u -1 new 1
Cerebellum parcellation...
    From con006//mni/con006_t1_n4_mni.nii.gz to con006//parc/con006_t1_n4_mni_seg.nii.gz
Post processing...
    From con006//parc/con006_t1_n4_mni_seg.nii.gz to con006//parc/con006_t1_n4_mni_seg_post.nii.gz
Transform back to original space...
    From con006//parc/con006_t1_n4_mni_seg_post.nii.gz to con006//con006_t1_n4_mni_seg_post_inverse.nii.gz
Report generation...
```

When ACAPULCO is finished, the following will be generated in the subject-specific folders for use in QC and group-level data aggregation (see below):
- <subject>_n4_mni_seg_post_inverse.nii.gz: parcellated cerebellum mask in original (subject space).
- <subject>_n4_mni_seg_post_volumes.csv: volumes (in mm3) for each of the 28 subunits generated by acapulco.
- PNG's (in 'pics' directory): sagittal, axial and coronal.

All other folders contain the outputs of the intermediate steps of the pipeline, and can be deleted after completing QC if desired.

**Batch this process for multiple images:**

A "for loop" can be used to process multiple images serially. Each image = ~5-6 minutes.

First, make a file called file_list.txt, which has the full path to each image on a separate line. An easy way to do this is to navigate to your 'input' directory, and type:

```
ls $PWD/*.nii.gz >> file_list.txt
```

This will need to be modified a little if your images are in subdirectories, eg:

```
ls $PWD/cont*/*.nii.gz >> file_list.txt
ls $PWD/patient*/*nii.gz >> file_list.txt
```

Move file_list.txt to your 'acapulco' directory, and change directory to your 'acapulco' directory.

```
mv file_list.txt ../acapulco
cd ../acapulco
```

Then, run the loop from inside your acapulco directory (copy-paste all of these lines into the terminal together):

```
for file in $(cat file_list.txt); do
```

```
subj=$(basename $file .nii.gz
mkdir output/${subj}
cp $file output/${subj}
singularity run --cleanenv -B $PWD:$PWD acapulco-0.2.1.sif -i
output/${subj}/${subj}.nii.gz -o output/${subj};
Done
```

Please be aware that the highlighted lines are supposed to be written together in one line.

***For advanced users and troubleshooting, see**:*
https://gitlab.com/shuohan/keras-unet-cerebellum/-/blob/master/README.md

## STEP 2: Estimating ICV with Freesurfer

The quality check scripts use cerebellar volumes corrected for Intracranial Volume (ICV) for the detection of statistical outliers. This approach is more sensitive to detecting 'true' outliers by removing variability that is due to head size, and may also be less prone to detecting false outliers.

**ICV is estimated using Freesurfer autorecon 1. It does NOT need to be re-run if you already have Freesurfer outputs (see below). If you do not need to re-run Freesurfer, please skip to page 8. Otherwise, please follow the protocol described below.**

### Required software:

Freesurfer 7 https://surfer.nmr.mgh.harvard.edu/fswiki/DownloadAndInstall

- For Windows, Linux version must be downloaded and installed into a directory of preference which will later be used as *FREESURFER_HOME*.

After the download, open the Ubuntu terminal.

### Directory Structure and set-up:

Create a 'freesurfer' directory in your 'enigma' directory (beside 'acapulco' and 'suit')

Every time you work with FreeSurfer, you must set the following variables:

```
Bash
export FREESURFER_HOME=<freesurfer_installation_directory>
source $FREESURFER_HOME/SetUpFreeSurfer.sh
```

- *In Linux, the installation directory is generally: /usr/local/freesurfer*
- *In Mac, the installation directory is generally: /Applications/freesurfer*

Replace <path> in the following:

```
export SUBJECTS_DIR=<path>/enigma/freesurfer
```

(or from within your 'freesurfer' directory: *export SUBJECTS_DIR=$PWD)*

For a single subject, from inside your freesurfer directory:

```
cd <path>/enigma/freesurfer
recon-all -i ../input/<subject>.nii.gz -s <subject> -autorecon1
```

- If you Ubuntu says it cannot find the licence:
  - Go to the website to register: https://surfer.nmr.mgh.harvard.edu/registration.html
  - After the registration, you will receive **licence.txt** into your mailbox with the email you provided for registration.
  - Put **licence.txt** inside the folder where FreeSurfer is installed.
  - After this step, you can run FreeSurfer fully functioning.

- If you face the error:
  You are trying to re-run an existing subject with (possibly)
  new input data (-i). If this is truly new input data, you should delete
  the subject folder and re-run, or specify a different subject name.
  If you are just continuing an analysis of an existing subject, then
  omit all -i flags.

  Go to the folder where FreeSurfer is installed, go to the subjects folder and delete the corresponding subject folder.

**Running ACAPULCO Quality Control**

3 steps:

- Visual quality check of the segmented images for each subject (html file with images for all subjects) 30 seconds per/subject
- Quantitative statistical outlier detection (criterion of +/- 2.698 SD from the mean for each cerebellar lobule (box plots) 30 seconds per/subject
- Detailed QC- those subjects flagged as outliers from steps 1 and 2 above. FSLeyes (or alternative images viewer software) to scroll through T1 and cerebellum mask images slice-by-slice

**Rstudio and Pandoc**

- Download and install “Rstudio” from https://www.rstudio.com/products/rstudio/download/
- Make sure you have pandoc in your directory within Rstudio installation directory RStudio/bin/pandoc. If not, you can manually download it from: https://pandoc.org/installing.html for Linux.
- Extract the downloaded zip file in your directory RStudio/bin
- Make sure that Sys.setenv variable is set to:
  Sys.setenv(RSTUDIO_PANDOC="<your Rstudio installation directory/bin/pandoc> ")

**Enigma Files**

- Download the “R.sif” singularity container and save to your “acapulco” directory. The container is used to run QC and ICV calculations.
- Download the QC scripts. Unzip and place the contents in /enigma/acapulco directory. You should see 3 scripts: “QC_Master.R”, “QC_Image_Merge.Rmd” and “QC_Plots.Rmd”

In terminal, write:

```
singularity exec --cleanenv -B /full/path/to/acapulco/output:/output -B
/full/path/to/Freesurfer:/freesurfer R.sif Rscript /path/to/QC_Master.R
/output /freesurfer
```

Where *full path to freesurfer* refers to the freesurfer folder you created in the enigma directory.

Ø Make sure the names of the subject specific folders in your acapulco output directory and freesurfer output directory. These will be concatenated by the subject folder names for the plots.

A folder called QC+Vols will be generated which contains the following:

1) “QC_images.html” file contains the ACAPULCO–generated QC images of all subjects in all 3 sections (coronal, sagittal, transverse).

2) “Plots_for_Outliers.html”, which contains both uncorrected and corrected plots for visualisation of the data distribution and statistical outlier detection. NOTE: We use the corrected plots for identification of statistical outliers.

3) “Outliers.Corrected.csv” and Outliers.csv file: these contain statistical outliers from the corrected and uncorrected cerebellar volumes.

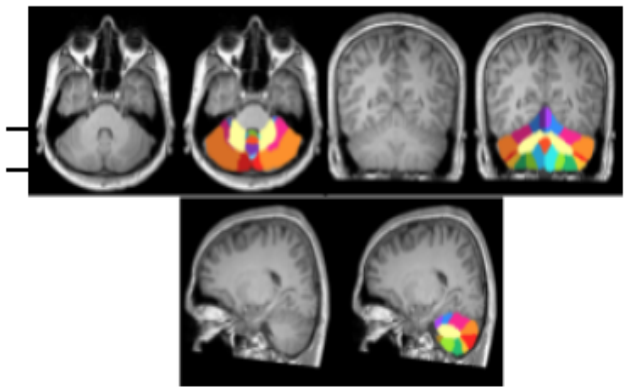
4) “Cerebel_vols.Corrected.csv” and “Cerebel_vols.csv” files containing the cerebellum volumes for each lobule, corrected and uncorrected for ICV, respectively.

**Examining the QC Images generated by ACAPULCO**

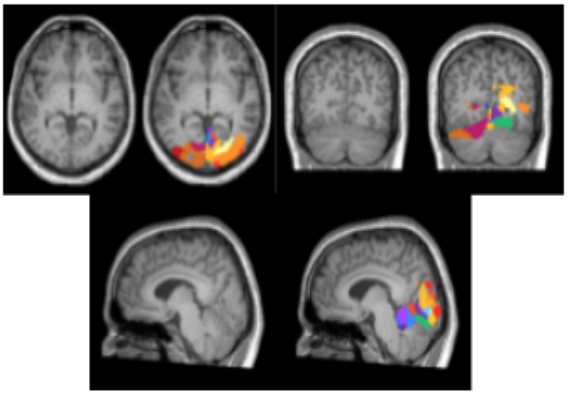
The file “QC_Images.html” contains the ACAPULCO–generated QC images of all subjects in all 3 sections (coronal, sagittal, transverse). Open in a web browser and quickly (~10 second per subject) scroll through the images to identify obvious failures or systematic issues. Note the subject IDs of failed or suspect segmentations for follow-up. Tip: holding down the command and F keys on a Mac, (control & F on PC), will bring up a search bar and allow you to enter in a subject ID to search. Command and + keys will zoom in.

The following examples come from the initial ENIGMA Cerebellum Volumetrics Pipeline Tutorial document by the Enigma - Ataxia Working Group.

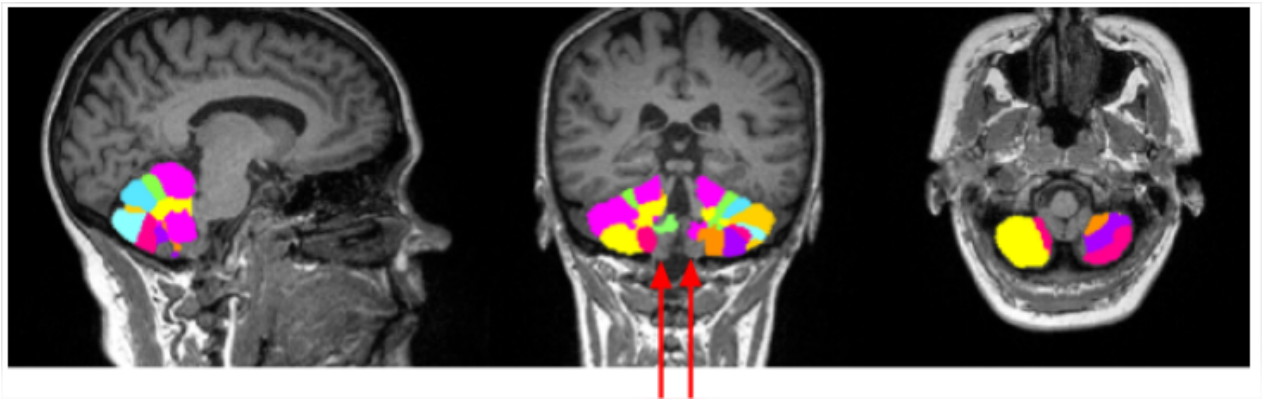
**Example of a good segmentation:**



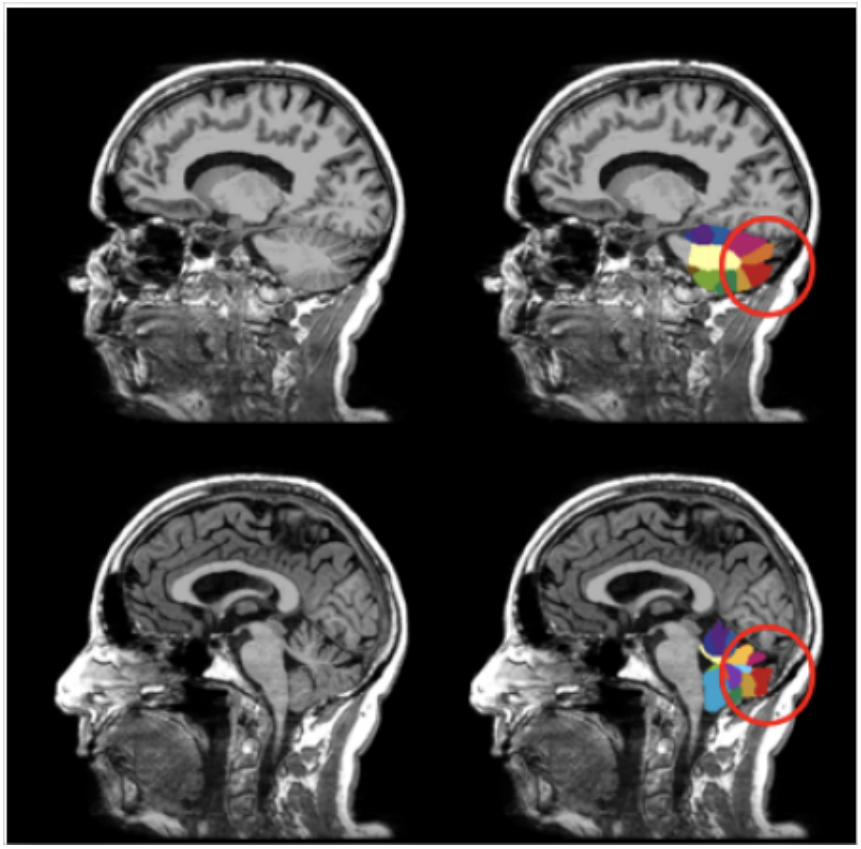
**Example of a global failure:**



**Examples of a more subtle mis-segmentation:**



Above: a mis-segmentation. Here, lobules VIIIa and IX have been completely missed. We would recommend excluding those lobules for this participant. Below: a mis-segmentation. Here, all of the data were within normal range (no statistical outliers were detected), despite an obvious underinclusion of Crus I and Crus II. We would recommend excluding Crus I and Crus II lobules for this participant.



**Examining the distribution of the volume estimates**

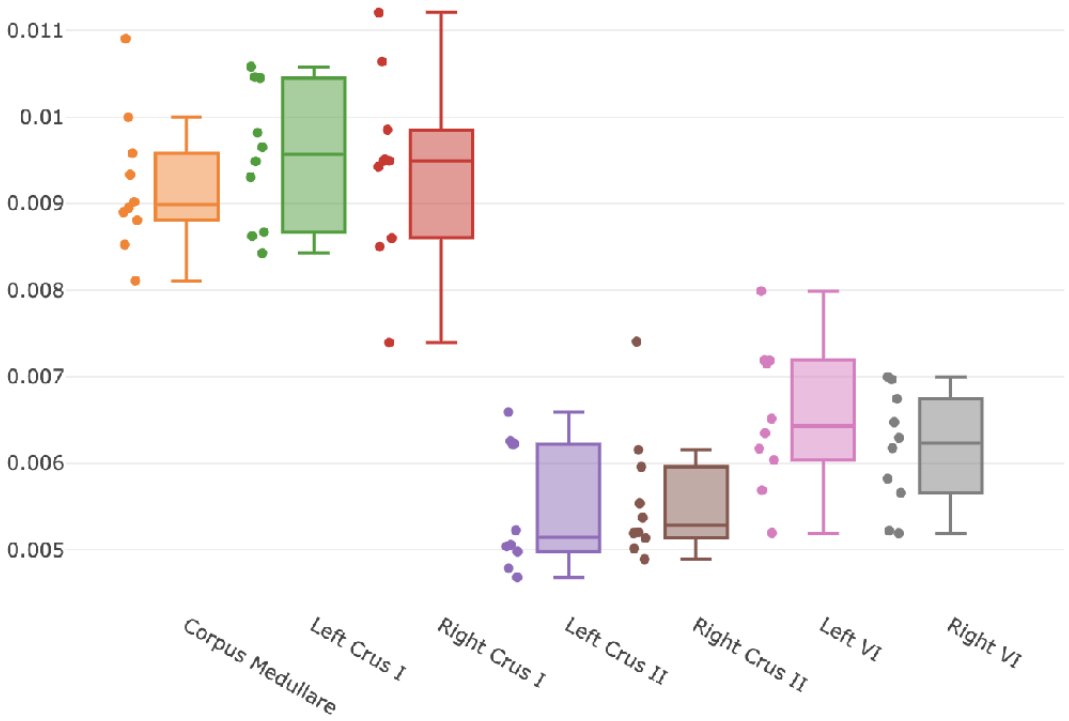
The file "**Plots_for_Outliers.html"** contains interactive plots to visualise the data distribution.

We have provided both uncorrected and corrected plots for the user to compare, however for QC purposes, please scroll down to the **"Corrected"** plots. Here, the y axis represents the volumes as a fraction of the total ICV and so will be very small values. Outliers will be data points above or below the whiskers of the boxplot. **Hovering** over those data-points should display the Subject ID. Also, you can **click and drag to zoom** over a particular graph. Outliers are also recorded in **Outliers.Corrected.csv**

# Corrected Volumes

## Total Cerebellar Volume


## Large Lobules



**Outlier confirmation and exclusion:**

Open the file **Outliers.Corrected.csv**.

For each segment and each subject, outliers (2.698 s.d above or below the mean) are denoted
by a '1'. The final column contains the total number of segments identified as outliers for each subject.
Each subject having one or more outliers should be manually inspected using FSLEyes (or another image viewer of your choice). A determination will need to be made whether to include the abnormal segment or not – i.e., is there a segmentation error, or is it just normal variability in the individual's anatomy? Each segment is considered individually, so a small number of segments can be excluded for an image, while the remainder are retained if correct.

Two decisions need to be made in each case:

1) *Do one or more segments need to be excluded from the final dataset?*
   - o If Yes (outlier is confirmed), exclude this segment(s) from the analysis by replacing the volume estimate with "NA" (no quotes) in the corresponding cell of the **Cereb_vols.csv** file (*not* the Cerebel_vols.Corrected.csv file)
   - o for that subject, and note the excluded segments in the quality check spreadsheet (**ENIGMA_Cerebellum_QC_spreadsheet**)
2) *Do segmentation errors result in some of the cerebellum being excluded from the mask?*
   - o If Yes, indicate 'No' in the **ENIGMA_Cerebellum_QC_spreadsheet**, in addition to the procedure in #1 for the affected segments. (The subjects will need to be excluded from analyses using the SUIT module).

The spreadsheet will look something like this:

| Quality Assurance- ACAPULCO | | | | | |
|---|---|---|---|---|---|
| **Subject_ID** | **Checked** | **Segmentation Notes e.g., list here individual lobules that have been under or overincluded, if they were detected as an outlier, if segmentation failed etc. Write "Exclude" followed by the lobule(s) to be excluded** | **Does mask incl whole cerebellu** | **Exclude?** | **Cohort** |
| Cont_01 | Y | L and R Crus I underinclusion | Yes | No | Melb |
| Cont_02 | Y | Good | Yes | No | Melb |
| Cont_03 | Y | Good | Yes | No | Melb |
| Cont_04 | Y | Good | Yes | No | Melb |
| Cont_05 | Y | Good | Yes | No | Melb |
| Cont_06 | Y | L lobule IX underinclusion | Yes | Exclude L lobule IX | Melb |
| Cont_07 | Y | Good | Yes | No | Melb |
| Cont_08 | Y | Good | Yes | No | Melb |
| Cont_09 | Y | Good | Yes | No | Melb |
| Cont_10 | Y | Not good segmentation. Some overinclusion of Left Cus I a | Yes | Exclude L Crus I and II | Melb |
| frda_01 | Y | Several lobules with segmentation error. Very scrappy alo | No | Yes- exclude subject | Melb |
| frda_02 | Y | Segmentation is good considering atrophy of cerebellum | Yes | No | Melb |
| frda_03 | Y | Minor overinclusion of R lobule VIII | Yes | No | Melb |
| frda_04 | Y | Good | Yes | No | Melb |
| frda_05 | Y | Good | Yes | No | Melb |
| sca2_01 | Y | Good | Yes | No | Melb |
| sca2_02 | Y | Dealt well with atrophied cerebellum | Yes | No | Melb |

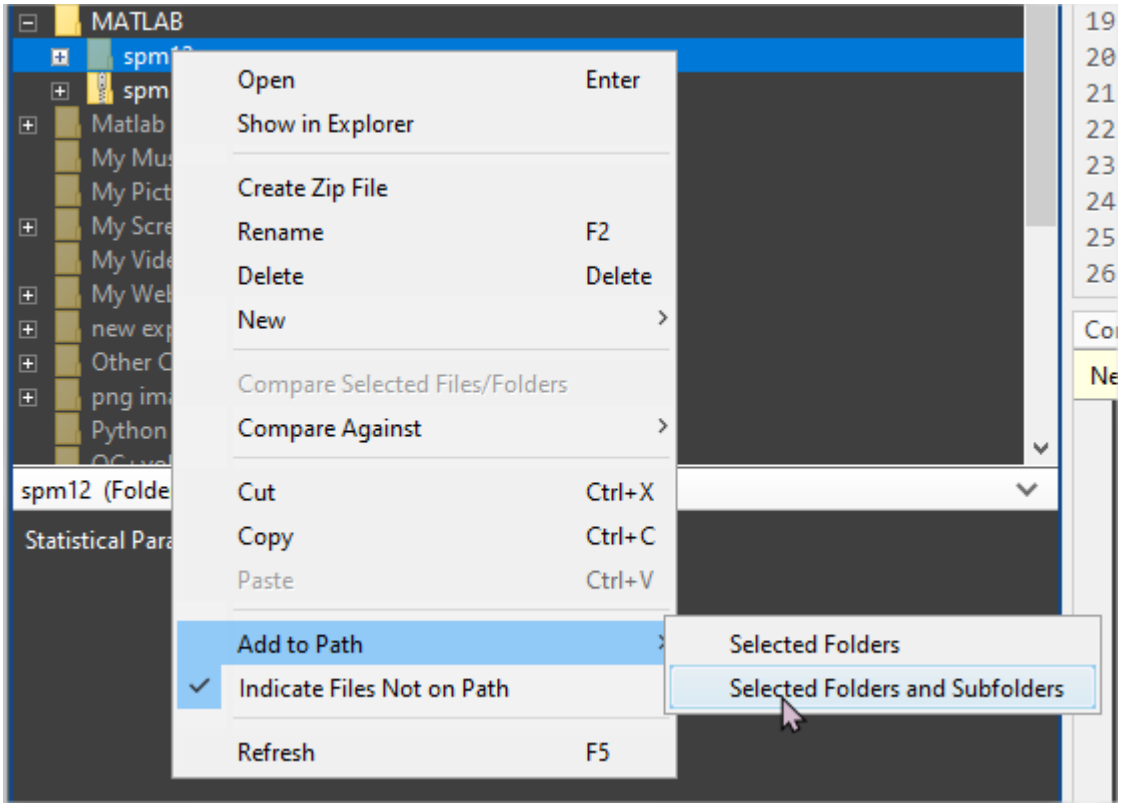
## SUIT

**Required Software**

1. MATLAB
2. SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/)
3. SUIT Toolbox, ver 3.4 (http://www.diedrichsenlab.org/imaging/suit_download.htm)
   - o Unzip the package in your 'spm12/toolbox/' directory.
4. ENIGMA SUIT Scripts (enigma_suit_scripts.zip)
   https://drive.google.com/drive/folders/1Wr-7tb0c4vZBGz_cQo1bGNQYW3XiqDsE
   Unzip and place in the 'spm12/toolbox/' directory

**CRITICAL:** This pipeline requires the ACAPULCO module to have already been run. The ACAPULCO outputs provide the inputs for this SUIT pipeline. Images in which the ACAPULCO mask does not include the whole cerebellum and must be excluded.

- Download and Install Matlab
- Download and Install SPM12 https://www.fil.ion.ucl.ac.uk/spm/software/download/
- Unzip spm12.zip in a folder of your choice, such as home/usr ...
- Start MATLAB and add SPM into your path, either using *File* > *Set Path* > *Add Folder...*




or typing

```
>> addpath /home/usr …
```

in MATLAB's workspace.

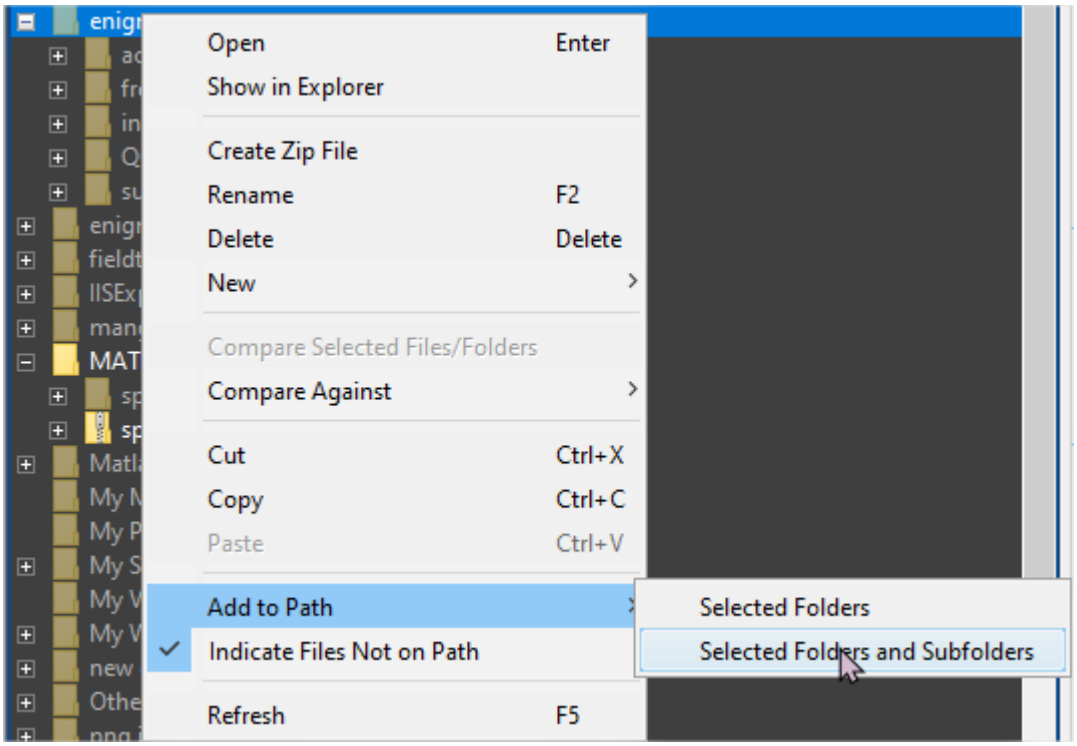
- Launch SPM by typing

```
>> spm
```

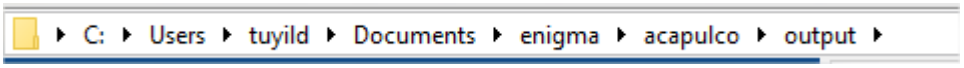
For more information, follow the link
https://en.wikibooks.org/wiki/SPM/Installation_on_Windows#Installation

- Create a new folder named “suit” in your enigma directory.
- Create a new folder named “enigma_suit” in the MATLAB directory, inside the toolbox folder and put “suit_enigma_all” in this new folder.
- In the MATLAB > Current folder, go to your enigma directory.

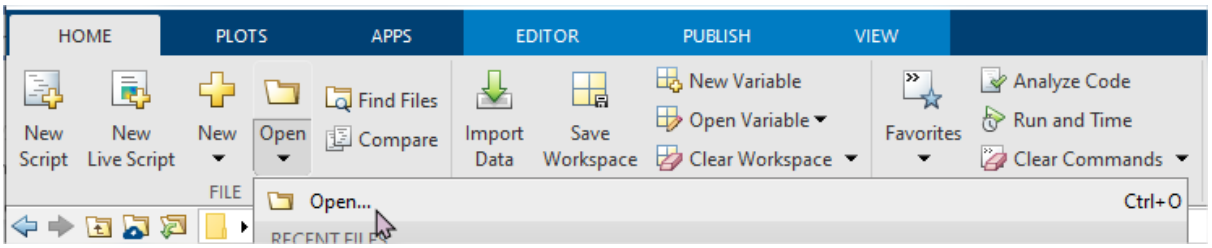

- Add all the enigma folders to path.




- In the current folder, click on enigma > acapulco > output. This puts you in the output directory. You should see on Matlab that you are in your output directory:

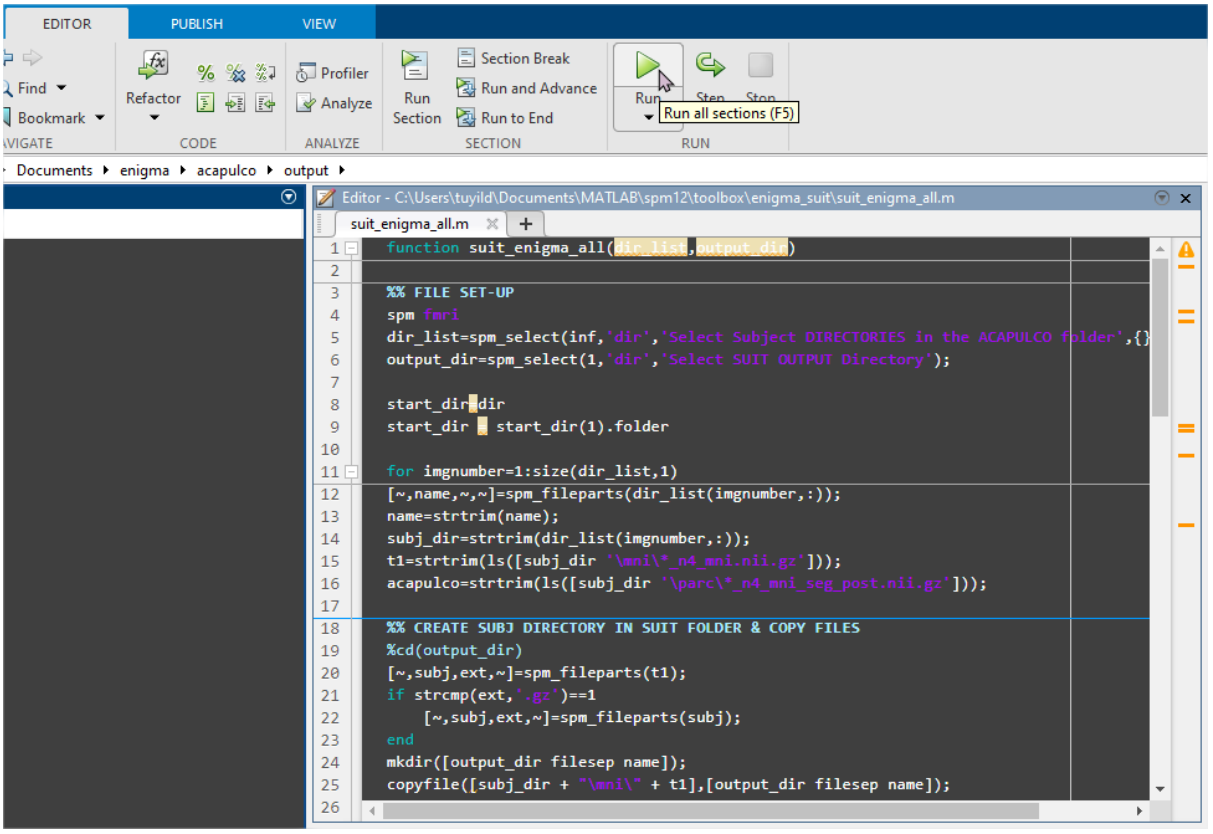



- In MATLAB, go to the section Home and open the script provided “suit_enigma_all” in MATLAB > toolbox > enigma_suit directory.

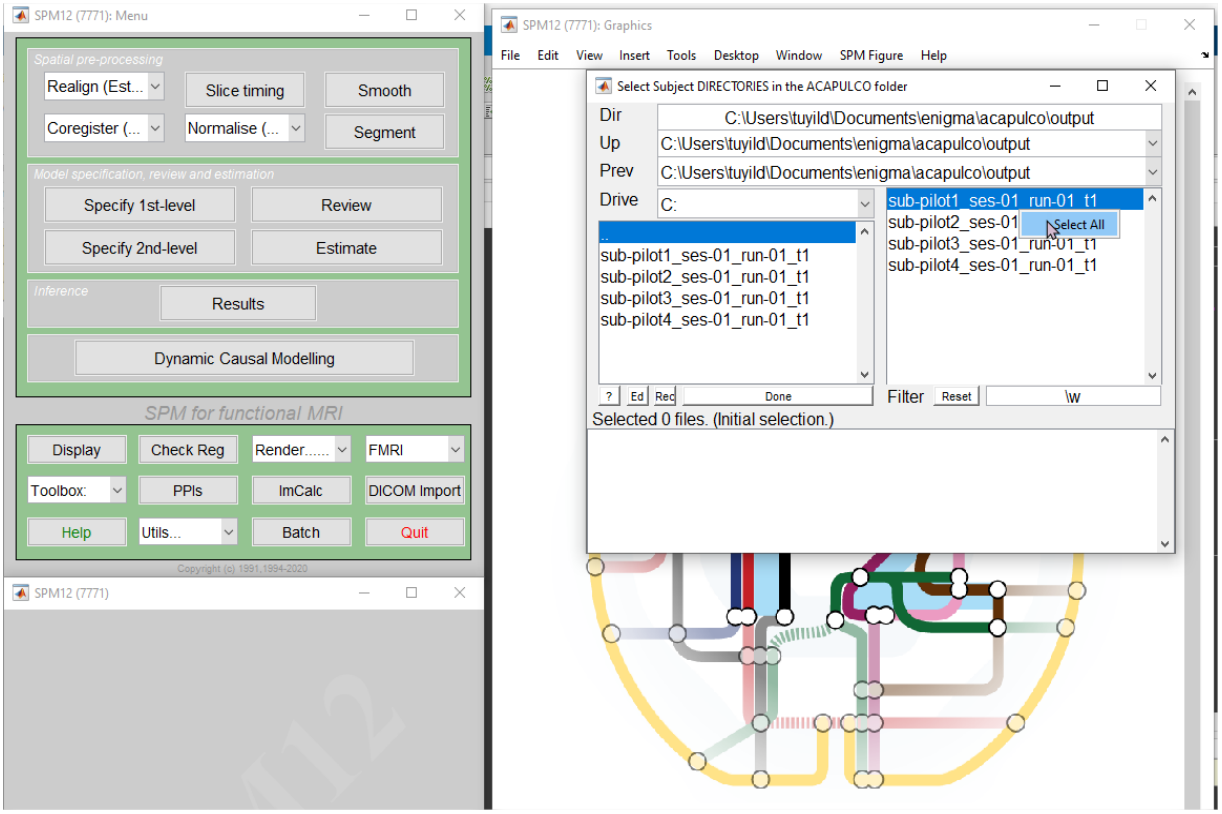



The next examples are taken from Matlab for Windows, but you will see your directories in Linux.

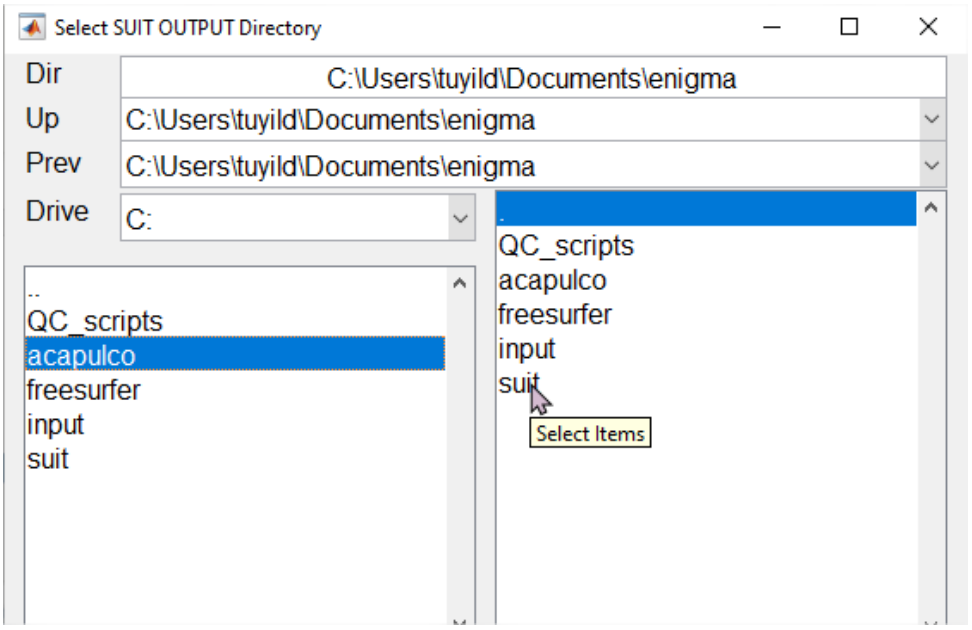
- In MATLAB, go to the section Editor and Run the script.




- This will start spm and ask you to select subject directories in the Acapulco folder.




- You can select all your acapulco output subject folders here.
- After this step, you will be asked to choose a Suit Output Directory. This is where the outputs of Suit will be put for each subject. Navigate through the enigma directory and choose the suit folder you created.




- Suit pipeline will start running after this step!

**OPTION 2: To run from the command line (for scripting/looping):**

From the MATLAB command line. The first input to the function is a subject-specific directory containing ACAPULCO outputs. The second input is the parent output directory for SUIT analyses.
Paths can be relative or absolute:

```
>>
suit_enigma_all('/path/to/acapulco/output/subjdir','/path/to/suitoutputdir')
```

OR from the TERMINAL (bash/shell) command line:

```
>> matlab -nodisplay -nosplash -r
"suit_enigma_all('/path/to/acapulco/output/subjdir','/path/to/suitoutputdir')
, exit
```

If your spm12 and/or enigma_suit directories are not permanently saved to your MATLAB path (e.g., when working on a shared cluster), you'll need to add that step to the TERMINAL command:

```
>> matlab -nodisplay -nosplash -r "addpath('/path/to/enigma_suit'),
suit_enigma_all('/path/to/acapulco/output/subjdir','/path/to/suitoutputdir'),
exit"
```

**Batch processing of multiple images from the command line:**

Create a text file called subject_list.txt, containing the ID of every subject in the ACAPULCO directory for which you want to run suit. Save the file in the ACAPULCO directory.
From the TERMINAL (bash/shell) command line:

```
>> export WORKING_DIR=/path/to/acapulcodir/output
export OUTPUT_DIR=/path/to/suitoutputdir
cd $WORKING_DIR
cat $SUBJECT_LIST | while read i; do
matlab -nodisplay -nosplash -r "suit_enigma_all('$i','${OUTPUT_DIR}'),exit"
done
```

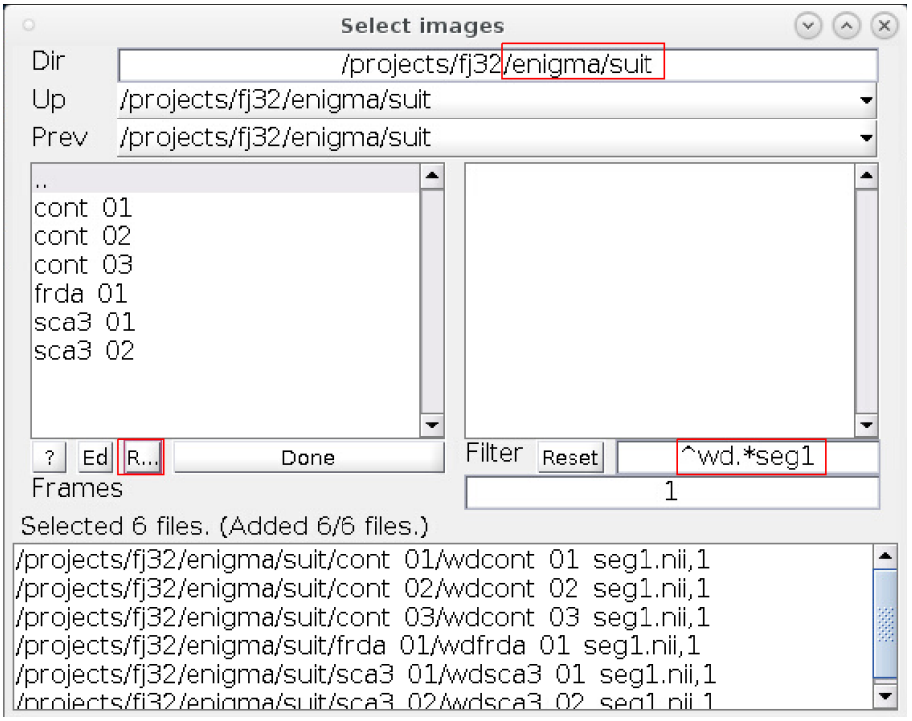
The script will:
- Copy the N4 bias corrected, MNI-aligned (rigid-body) T1 image, and the ACAPULCO cerebellum mask into the output directory
- Segment the grey and white matter of the cerebellum
- Correct for overinclusion errors in the segmentation using the ACAPULCO mask

- DARTEL normalise and reslice the data into SUIT space, with Jacobian modulation so that the value of each voxel is proportional to its original volume.

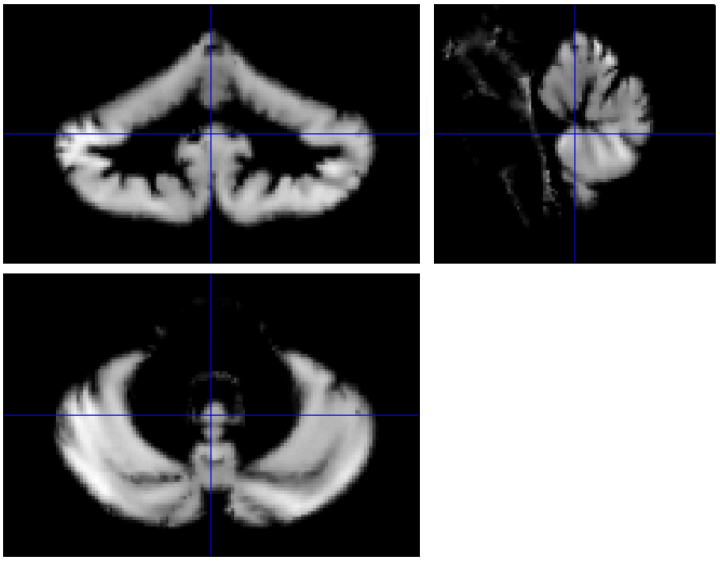
**Outlier Detection and Quality Control:**

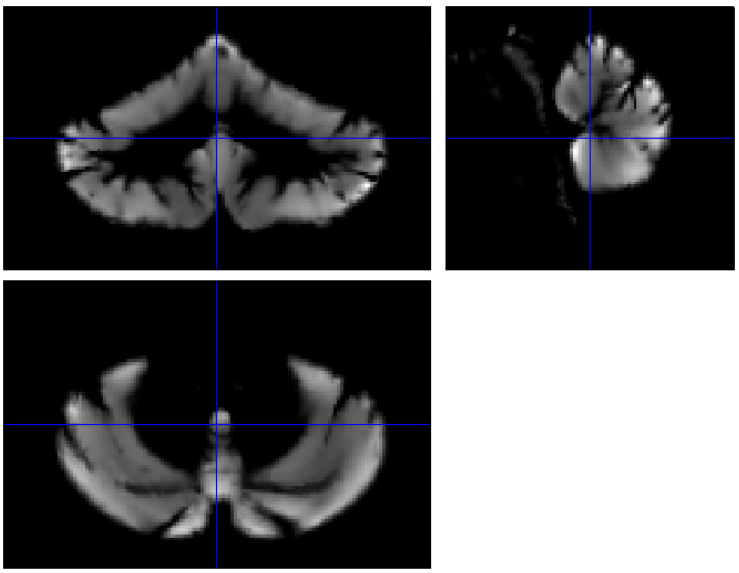
**1.** Visually inspect the normalised, modulated images (wd*) for major failures. In MATLAB, type the command: **spm_display_4D**

Manually select the “wd*seg1” images from the suit subfolders, or navigate to your ‘suit’ directory, insert “^wd.*seg1” in the Filter box (no quotations) and press ‘Rec’ button. Press Done.



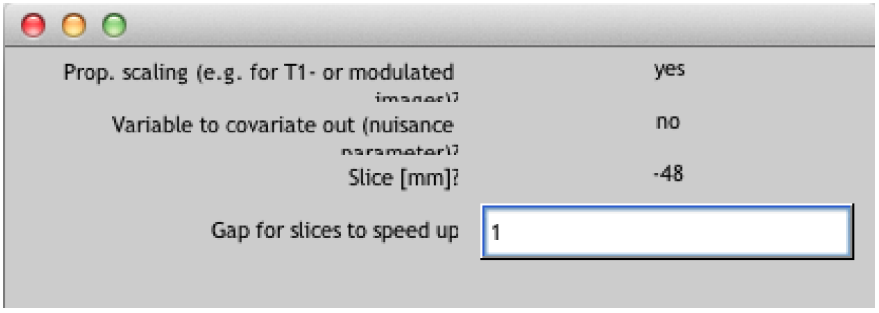

Scroll through the images to ensure they are all well-aligned. Correctly normalised images from a healthy control (left) and an individual with a heavily atrophic cerebellum (right) are shown below. Note that at this stage the between-subject anatomy is very similar (as they have been registered to the same template) and volume differences are instead encoded by differing voxel intensities.



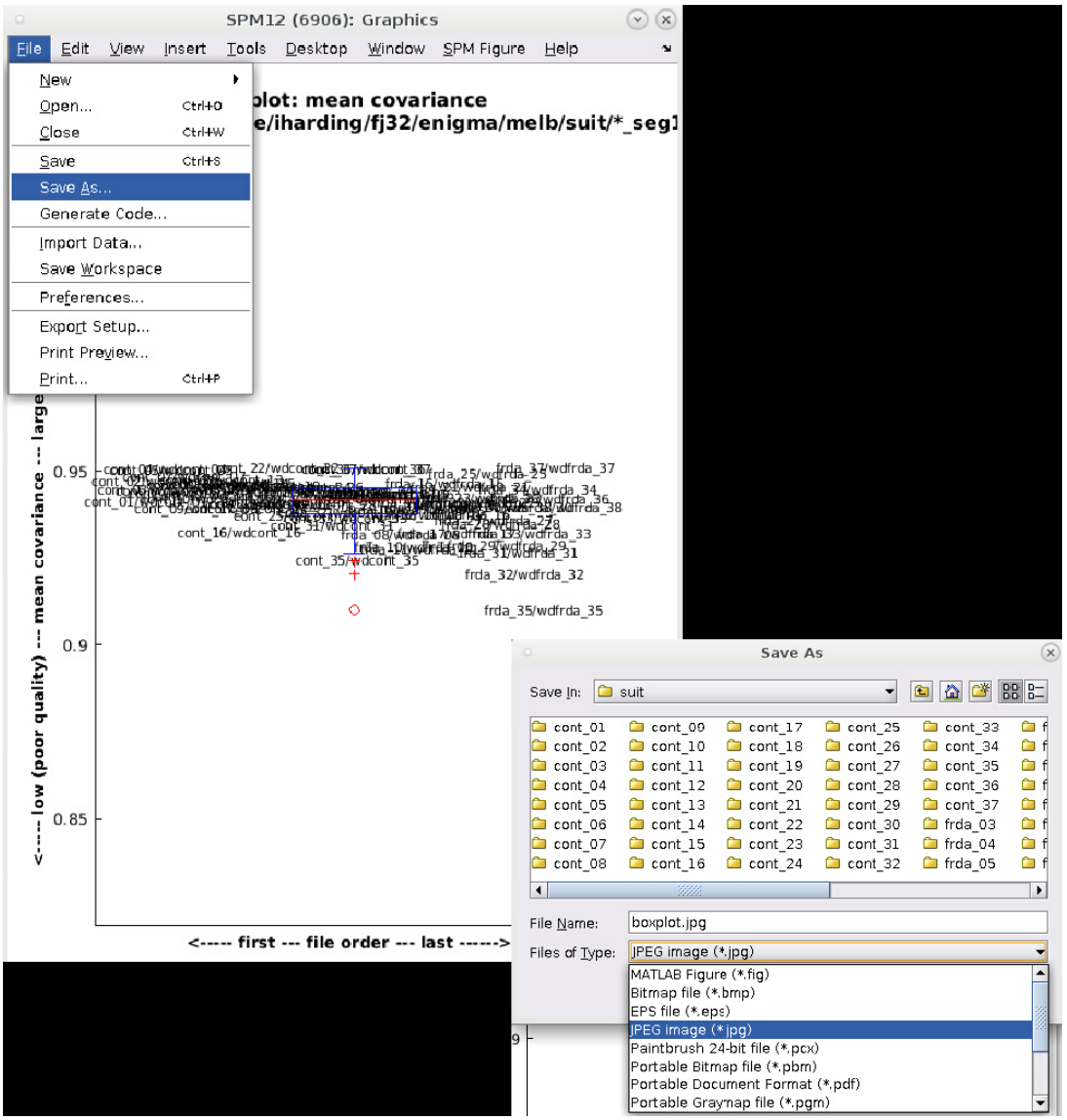


Major failures will be very obvious: blank images, large areas of missing tissue, unusual intensity gradients (i.e., bright voxels all at the top, dark voxels all at the bottom), etc. These images should be excluded from subsequent steps.

**2.** Check spatial covariance for outliers. In MATLAB, type the command:
**check_spatial_cov**
Select the “wd*seg1” images as per the previous step. Select the following options when prompted:




There will be several outputs you can explore, but the important output is the boxplot, which will display the mean spatial covariance of each image relative to all others in your sample:



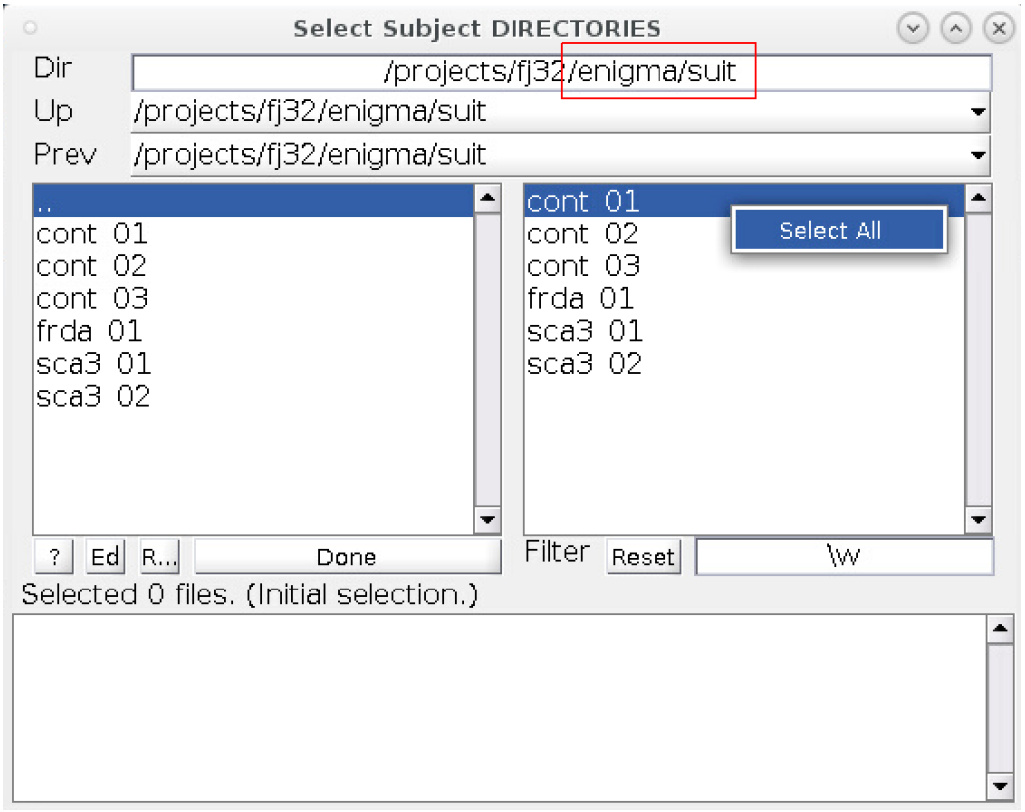

Data points that are >2s.d. below the mean will be identified in the MATLAB command window and will fall below the lower arm of the boxplot. For these, the <subj>_n4_mni.nii.gz

image in the SUIT folder should be inspected for artefacts (motion, anatomical abnormalities, etc.), image quality issues, or preprocessing errors.

If you are confident that the image quality and preprocessing is fine, and visual inspection of the modulated images in the previous step does not indicate an issue with segmentation & normalisation, retain this data in the sample. Otherwise, this data should be excluded.
Save the boxplot of the final sample in your 'suit' directory as a JPEG file: boxplot.jpg

**Packaging the Data for Submission:**

In MATLAB, type the command: **PacknSendSUIT**
Select the subject directories from your 'suit' directory that are included in your final cohort. e.g., Right click and 'Select All', then remove excluded subjects (i.e., outliers) by selecting them individually in the 'Selected Files' list at the bottom. Press Done.




This step will collate and zip all of the final processed images (the "wd" files) from the directory of each subject you select to be in your cohort, (plus the boxplot.jpg files if created in previous step). This archive will be labelled "SUIT_Final_All.tar" and written to your 'suit' directory.

An archive labelled "SUIT_Final_All.tar" will be written to your 'suit' directory.