PANTHEON
Stroman Lab, Queen's University
stromanp@queensu.ca

May 28, 2024

PANTHEON is a python software package for the analysis of functional MRI data. It was developed to be able to analyze data from the entire central nervous system (CNS), including all levels of the spinal cord, brainstem, and brain.

This software is for research purposes only, and it is a work-in-progress. No guarantees are given about the reliability or sensitivity of the results. It is up to the user to understand what they are doing, and the quality of the results will depend on the quality of the data that are analyzed. It is necessary for the user to understand how MRI data are acquired, what the data show, and how the data are organized. Detailed background information about MRI can be found here: https://www.routledge.com/Essentials- of-Functional-MRI/Stroman/p/book/9781439818787#

The current version of Pantheon was developed in Windows, and has been adapted for use on Mac operating systems. The only noted issues have been with the display of elements in the graphical user interface. It has not been tested on Unix systems.

The individual components of the software can be run from the command-line, but it is intended to be run primarily using a graphical user interface (GUI) that is provided. The user interfaces and programs are designed to support processing and analyses of large sets of data.

Notes about the data that can be analyzed:
The data must be acquired in a way to enable some analysis steps to function, such as spatial normalization. This means the data must include enough distinct anatomical features to identify the region, and so studies of the thoracic spinal cord, for example, can be difficult to normalize without additional position information input by the user. For most studies of the brain, brainstem, cervical cord, or lumbar cord regions, the spatial normalization appears to work well.

The key elements of PATHEON:

1) organization of a database structure for identifying which data set(s) to process/analyze
2) conversion from DICOM format to NIfTI format
3) computation of spatial normalization parameters
4) pre-processing including a) co-registration (motion correction), b) slice-timing correction, c) applying spatial normalization to each volume, d) spatial smoothing, e) creation of models of predicted BOLD responses and modeling of physiological noise and bulk motion, f) "cleaning" of data by fitting and removing physiological noise and bulk motion effects
5) general linear model (GLM) fitting of predicted BOLD responses to fMRI data
6) definition of regions-of-interest and clustering within regions
7) connectivity analyses by means of temporal correlations, structural equation modeling (SEM), and structural and physiological modeling (SAPM)
8) group-level analyses to identify consistent features of across groups, or features that are correlated with covariates (participant characteristics, behavioral data, etc.)
9) visualization of connectivity analysis results

1. INSTALLATION

This is the recommended method but other configurations are possible, such as different user interfaces for python.

STEP 1:
Use Python 3 by installing Anaconda from https://www.anaconda.com/distribution/

STEP 2:
Install Pycharm. This is the programming environment program:
https://www.jetbrains.com/help/pycharm/installation-guide.html

STEP 3 (option 1):
Create new project from VCS (version control)
Fill in information to select pantheon-fMRI from GitHub repository

Configure the project:
Set "Preferences" ➔ in left column under "Project: pantheon-fMRI, select Python Interpreter Beside the "Python Interpreter" box, select the gear wheel icon and select "add"
Select "Conda Environment" and set the version of python to run
Finish making selections to create the Conda environment and return to the Python Interpreter page The installed packages are listed on this page, select the Conda icon to the right of the "+"
Select "+" to add packages:
-install packages:
dicom2nifti
numpy
matplotlib (version 3.2.2 for python 3.8, later versions seem fine for higher python versions)
nibabel
openpyxl
pandas
Pillow
scikit-learn
xlrd (choose version 1.2.0 for python 3.8)
dipy
statsmodels

STEP 3 (option 2):
(may need extra steps with this method to be able to update the project automatically using Git)

a) Configure Pycharm Create a new project
…from the options available when setting up the project:
- choose new environment using Virtualenv
- base interpreter – python 3.8, 3.9, 310 all seem to work fine

- do NOT choose to inherit global site packages (will configure these)

b) Get pantheon-fMRI from GitHub by downloading a zipped copy
- copy the python files and "template" directory into the "venv" folder that was created for the project

c) Setup the environment in PyCharm for patheon
- go into "Settings" for the project
-install packages (same as above for the other setup options):
dicom2nifti
numpy
matplotlib (version 3.2.2 for python 3.8, later versions seem fine for higher python versions)
nibabel
openpyxl
pandas
Pillow
scikit-learn
xlrd (choose version 1.2.0 for python 3.8)
dipy
statsmodels

STEP 4

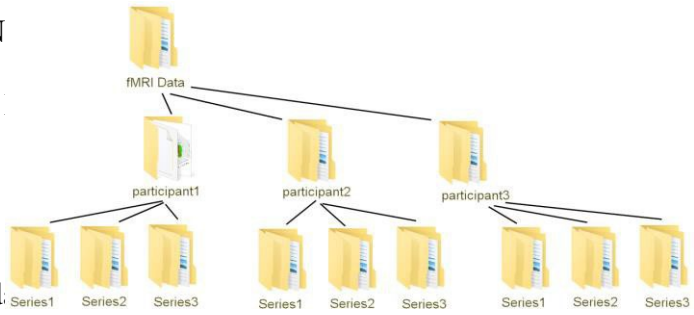
run pantheon.py in Pycharm

2. SETTING UP A DATABASE and ORGANIZING YOUR DATA

Organizing your data

The data to be analyzed can begin in DICOM format, or can already be converted to NIfTI. If they are already in NIfTI format the images must be oriented to match the MNI template, with the three spatial dimensions x,y,z corresponding to left➔right, posterior➔anterior, and foot➔head or inferior➔superior, respectively (i.e. "RAS" orientation)

If starting with DICOM data, all of the DICOM data from an imaging session (i.e. multiple series from the same person) can be in the same folder initially, but they will be organized into one series per folder during the conversion process, as described below. After the DICOM data are organized into one series per folder, the folder must contain only the image data for the series, with no other data or sub-folders. This is for the process of converting multiple DICOM files for one series into a single N



It is preferable, for ease of organizing your data, to within one main folder.

Defining a database file

The information needed to help the Pantheon software find and organize your data is kept in a "d database file easy to modify, view, etc.

The database file is a key element of doing automated analyses across multiple data sets, group analyses, etc. and for specifying covariates such as participant characteristics and behavioral data. An example of a database file is included with the Pantheon repository on GitHub.

The database file is an Excel file that contains a sheet named "datarecord". Additional sheets will be used as described later, but only one should be named "datarecord".
The datarecord sheet must contain the following columns: a number column, as well as **patientid**, **studygroup**, **pname**, **seriesnumber**, **niftiname**, **TR**, **normdataname**, **normtemplatename**, **paradigms**, **datadir**.
If you want to use a reference image to guide brain normalization, it is necessary to also include a column with the title **norm_bridge_ref** .
Each line in the database file will provided information for one series in your data set. As a result, you will likely have multiple lines for each participant (one for each series).

Example database file:

	patient id	study group	pname	series number	nifti name	TR	normdata name	normtemplate name	paradigms	datadir
0	pid1	sg1	pname1	series num1	tbd	TR	tbd	ccbs	paradigm1	datadir
1	pid2	sg2	pname	series num2	tbd	TR	tbd	ccbs	paradig	datad

			e2						m1	ir
2										
…	…	…	…	…	…	…	…	…	…	…
…										

The columns can be in any order, but the column names must be as specified, and they are case sensitive. You can add any additional columns that you want, such as the participant's age, sex, etc., or behavioral data such as performance metrics on a task, pain ratings, etc. These additional columns can be used as covariates in later analyses.

Defining the "datarecord" sheet in the database file

Column title	Description	Examples
datadir	This the full path to the top-level folder that contains your data folders. It could be different for different participants or groups, but for convenience you might want to make this the same for all of your data Note: Do NOT include a file separator (\ or /) at the end of the datadir name	C:\\study1\study_data
patientid	this can be any identifier assigned to your participant	participant1; patient_ABC; etc.
studygroup	this can be used to identify a participant group, or a study condition, etc.	healthy; patient; control_group; treatment; etc.
pname	This is the full folder name beginning from the folder specified under "datadir" The full path name to the folder will be used as datadir+pname (i.e. names concatenated together with the appropriate file separator) Note: Do NOT include a file separator (\ or /) at the end of pname. With the examples given, the full resulting path name to your data would be taken as C:\\study1\study_data\ controlgroup\participant1	controlgroup\participant 1
seriesnumber	This is the series number, as determined by the MRI system at the time that the data were acquired. This number is used to identify the data series that you want to anlyze	5; 6; etc.

niftiname	This is the path name of your NIfTI format data, relative to datadir. That is, it will include the NIfTI file name as well as the portion of the path included in pname. If you are starting with NIfTI data (i.e. not converting from DICOM) you can enter the file name here. If you are starting with DICOM data, you can enter "TBD" or any other term, and the value will be filled in by Pantheon when the data are converted from DICOM to NIfTI	If you have not converted the data to NIfTI yet: "TBD" or "undetermined" etc. If you have converted the data and know the name: controlgroup\participant1\Series5.nii
TR	The repetition time used when your data were collected.	3.0
normdataname	The name of the file that will contain the information needed to normalize your data. This will be filled in by Pantheon when the normalization parameters are computed.	"TBD" or "undetermined" etc.
normtemplatename	The region of the template to use for normalizing your data. For brain data, enter "brain" For data spanning the brainstem and spinal cord, enter "ccbs" For data spanning a range of the spinal cord enter the range as the upper cord segment "to" the lower cord segment, such as "T12toL5". The templates that are used, and how to use alternative templates, are described in later sections.	ccbs; brain; C5toL10; etc…
paradigms	This is the name of the sheet in the database excel file that contains the definition of your fMRI paradigm(s) that you wish to use for data analysis	paradigm1; any name as long as there is a sheet to match it in the excel file (described below)
norm_bridge_ref	[optional] This is the database number (in this same database file) that refers to the image series that should be used to guide the spatial normalization. For example, a T2*-weighted set of images might be acquired without using EPI as an undistorted reference. The EPI data can be normalized to match this reference scan, then the reference scan can be normalized to match the brain MNI template. The complete normalization mapping from the EPI data to the brain MNI template can then be determined.	5; 6; etc.

Defining the paradigm

You can define any number of paradigms to use for analyzing your data, such as different paradigms for different study conditions, etc.
The paradigm sheet name is listed under the heading “paradigms” as described in the table above. There must be a sheet with this name in the database excel file.

Definition of a paradigm

	dt	paradigms
0	1	0
1	1	0
2	1	0
…	…	…
20	1	1
21	1	1
22	1	1
…	…	…
40	1	0
41	1	0
…	…	…

The column “dt” indicates the time span for each entry, in seconds. The column “paradigms” indicates the timing of the stimulus, in terms of if the metabolic demand is expected to be lower or higher. The title of the column does not need to be “paradigms” this is just an example, any title can be used as long as it is not “dt”. Every column that is not the initial number column or the “dt” column is taken to define a paradigm. Note that the ellipsis “…” are meant to indicate missing rows, these would not occur in the actual paradigm definition but rather each successive number would be defined. When the basis sets are defined by Pantheon, a corresponding sheet will be created, with the suffix “_BOLD” added to the sheet name, and it will contain the paradigm definition after it has been convolved with the canonical hemodynamic response function (HRF).
The choice of “dt” should be sufficient to accurately describe your paradigm, it does not need to correspond with your TR value. Values less than 0.5 seconds are likely more detailed than necessary, and values longer than 5 seconds are not advisable because the HRF must be convolved with this paradigm definition, and some temporal detail is useful.
The example shown above describes 20 seconds when the metabolic demand is expected to be lower (such as a “rest” condition), followed by 20 seconds when the metabolic demand is expected to be higher (such as a stimulation condition), following by another period with lower metabolic demand.

3. Navigating around in Pantheon

Settings file

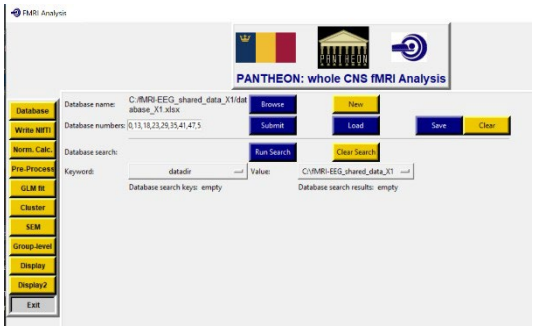
Pantheon keeps track of your settings so that when you restart the software your previous settings are kept, as much as possible. These settings are kept in a file named “base_settings_file.npy” in the “venv” folder of your pantheon installation. **If something goes wrong, and these settings become corrupted or cause some problem, simply delete the “base_settings_file.npy” file.** A new settings file will be created when you restart Pantheon.

“Pages” in Pantheon

“Pages” refers to the different views that you can have in Pantheon. The buttons on the left side of each page take you to other pages, with other options and steps for preprocessing and analysis.

Each page has a set of buttons to click on to initiate actions, or entry boxes for entering information. Some entry boxes have a “browse” button for selecting file names. All entry boxes have an associated “submit” button. When you enter a value into a box it is necessary to hit the “submit” button for the value to be set (unless you have used the “browse” button, in which case the value you pick is already set).

4. Indicating which data to work on with Pantheon – the "Database" page



The first page you see when you run Pantheon will look something like the figure on the right. If you are using Mac or Linux the colours might be different.

This is the "Database" page.
Here you can specify the database file to use. This is the excel file that is described above.
You can also specify the database numbers to use. These are the entries in the database. The numbers you enter correspond to the numbers in the first column of the database file, not the row numbers. You can enter numbers separated by commas, or as ranges, with the first and last number in the range separated by a colon (i.e. 0:10). You can also enter multiple ranges separated by commas. If you enter the word "all" and hit the submit button, every entry in the database file will be listed.

You can save a list of database entry numbers, or load a previously defined list using the "Save" and "Load" buttons. You can clear the numbers with the "Clear" button.
If you want to select a set of data based on features such as a particular study group, or participant characteristics (such as "male" or "female" if you have entered these) then you can select the column in the database file with the "Keyword" pull-down menu. The set of choices that you have in that column are then listed in the "Value" pull-down menu, and you can pick your search value from this menu. When you have made your choice, select the "Run Search" button, and the corresponding entries in the database will be listed in the "Database numbers" entry box, and you will be prompted to specify a name for saving the list.

You can choose a value for multiple Keyword choices, and you can also clear the set of Keywords using the "Clear Search" button.
The database entry numbers that are listed in the "Database numbers" entry box show the data sets that will be used in subsequent steps of pre-processing and data analysis.

Converting data to NIfTI – the "Write NIfTI" page
If your DICOM data files are not already separated into one series per folder, then you can select the "Organize Data" button to create subfolders and organize your data for you.
After your data are organized, click the "Convert" button to convert your DICOM files into a single NIfTI file for each series. The "Base name" indicated above the buttons indicates the root word that will be used to name your NIfTI files. For example, if you leave this choice at the default of "Series" then your data from series 7, for example, will be written to a NIfTI file named Series7.nii. After your data are converted, the niftiname entry in your database file will be updated to keep a record of the NIfTI file names.

5. Organizing data files and coverting to NIfTI – “Write NIfTI” page

The data that are generated by the MRI system is likely to be in DICOM format. Most data analysis packages have adopted the standardized NIfTI format both for each of managing the data and standardization across methods.
The steps on this page are to organize the data sets that are indicated by the database numbers entered on the “Database” page. Unless noted otherwise, all analysis steps here and on later pages refer back to the data specified on the Database page.

“Base name” refers to a text string that will be used as a portion of the resulting file name when the data are converted to NIfTI format. The name will also have the series number appended to it. This text string can be entered manually and then the “Submit” button must be pressed.

“Organize Data” carries out the step of sorting DICOM format data into folders named according to the series number, with one data series only in each folder. If the data are already organized then this step is not necessary.

“Convert” collects the data from each series and converts it into a NIfTI file.

6. Calculating normalization parameters – "Norm. Calc." page

It is necessary to calculate how to spatially normalize each data set before it can be applied to each volume during the pre-processing steps described below. The image templates that are used for normalization (and the corresponding region maps and white matter maps) are in the folders "venv\templates" and "venv\braintemplates" in the Pantheon package. The template that will be used is determined by the entry "normtemplatename" in your database file. Enter "brain" for brain data, "ccbs" for cervical cord and brainstem regions, or a spinal cord segment range as the upper segment "to" the lower segment, such as "C5toL10" or "T10toS5" etc. Note that normalizing data without any distinct landmarks, such as the brainstem, cervical enlargement, or lumbar enlargement, can be very challenging, and may not be accurate.

The brain templates are the MNI152 templates taken from Statistical Parametric Mapping (SPM12) (1) and the CONN15e software package (2), and the cord templates are a combination of the PAM50 template (3) and templates created in the Stroman Lab (4-8). Corresponding anatomical region-of-interest maps have also been defined from multiple sources, as described previously (2,9-19) (https://identifiers.org/neurovault.collection:3145 , www.med.harvard.edu/AANLIB/). (9,10,20-22).

The brainstem and spinal cord templates are loaded as sections depending on the region of interest, whereas for brain regions the entire brain templates are used. If you want to use other templates and white matter maps etc., you can substitute the files in these folders, but you need to use the same file names so that the software can find the templates to read.

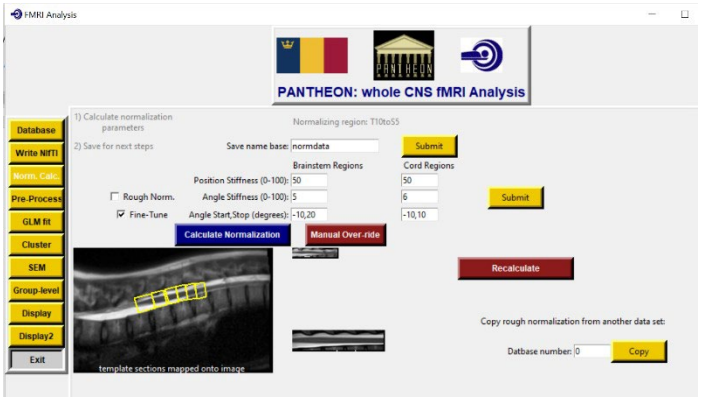
The page that is displayed will look different for normalizing brain regions compared to brainstem and spinal cord regions, because different user inputs are required, and different methods are used.
For brain regions, the normalization process uses the python package "dipy" (https://dipy.org/documentation/1.5.0/documentation/) which was developed for analyzing DTI data, but has useful functions that can be generally applied. As described in the documentation for DIPY, the normalization procedure in this method is based on the ANTs (Advanced Normalization Tools software: http://www.ncbi.nlm.nih.gov/pubmed/24879923, http://www.ncbi.nlm.nih.gov/pubmed/24817849)

For brainstem and cord regions, the normalization process consists of mapping sections of the template to the image data for one volume of your data series. This is done with predefined regions depending on the region of interest (normtemplatename) that you have specified. Once initial sections are mapped to distinct regions (such as the brainstem, or lumbar enlargement), they are checked for consistency, and then the sections are extended along the cord as a connected "train" to ensure that the length of the cord anatomy is not altered. The normalization procedure is based on the premise that the cord anatomy is likely much more consistent in size across people, than is the spine anatomy, and so vertebral features cannot be used as anatomical references, and the distance along the cord from specific features must be maintained as in the original image data.

The page includes information messages such as near the top/middle the region to be normalized is indicated. If this does not match what you expect, make sure you have correctly entered your database name and database numbers on the first page.

Input parameters for normalizing brainstem and spinal cord regions

"Save name base" indicates the name that will be given to the normalization data file, plus a suffix indicating the series number.

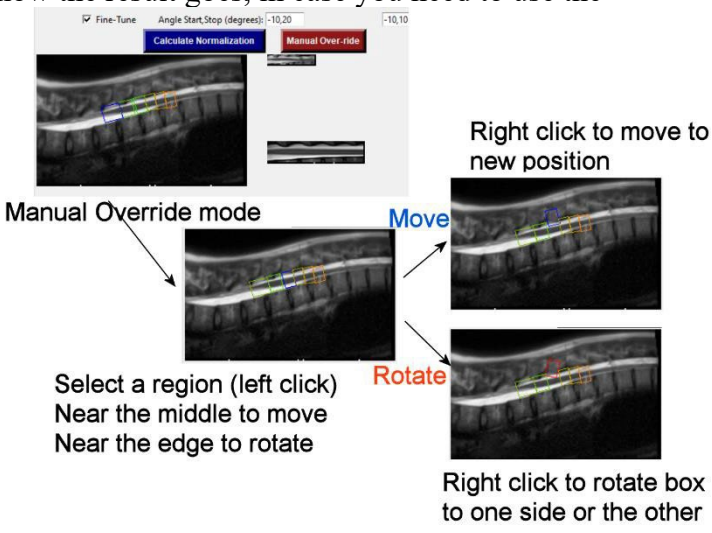


"Position Stiffness": This value is a relative indicator (0-100) for how much to allow the position and angle of a section to vary from the predicted position based on previous sections and their positions in the template. A lower value allows for more variation whereas a higher value pushes the result to be a closer match to the expected angle/position for each section.There are two values to allow for more flexibility in cord regions. The labels indicate "brainstem" and "cord" but for lumbar/thoracic cord regions the
lumbar sections, having more distinct anatomical landmarks, will be determined first and the remaining sections would be determined sequentially working up the cord, so in this case the first value would refer to the lumbar segments and the second value would still refer to the cord sections.

"Angle Stiffness": this is similar to position stiffness as it refers to how much flexibility is allowed between the estimated angle for each section and the expected angle for each section.

"Angle Start/Stop": this is the range of angles that will be searched for each section, in degrees. If you find that the sections are not fitting the cord/brainstem anatomy as they are not reaching a large enough angle from horizontal, then you can increase this range. However, a larger search range takes longer, because there are more angles to check.

"Rough Norm." and "Fine-Tune" checkboxes:
Typically you would select both of these boxes. "Rough Norm." indicates to apply the method of matching sections of the template to the data, and estimating how to normalize the data. The "Fine Tuning" is applied as a second step, to improve the fit of the rough-normalized data to the template. You may choose to do Rough Norm. first, and see how the result goes, in case you need to use the "Manual Override" option described below. After Manual Override, then you only need to apply the fine-tuning step.



Once you have made your selection of parameters, then hit the "Calculate Normalization" button to run the normalization. If you have selected multiple database numbers, this normalization will be applied sequentially to all of the data sets that are indicated.

Spatial normalization of brainstem and cord regions can be challenging because of variations in the anatomy, and because there is a large amount of surrounding anatomy that must be ignored by the normalization process. While the normalization process will work very well in most cases, it is occasionally necessary to correct errors in the process. This can be done with the "Manual Over-ride" button.

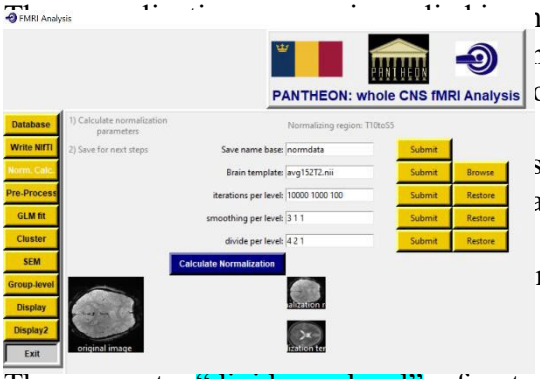
After selecting "Manual Over-ride" and right- clicking in the display window, the boxes outlining each section will be colour-coded to show if the sections have been positioned in the correct order along the cord. Even in extreme cases the boxes will probably be in the correct order. Right clicking near the middle of a box will make it turn blue, indicating it is in "move" mode, and clicking near the edge of the box will make it turn red, indicating it is in "rotate" mode. To move the box, right- click at the new position where want the box to be centered. To rotate the box, right-click on one side or the other to rotate the box in that direction. Make adjustments until the box is positioned where you want. Remember, it does not need to be perfect, the fine-tuning stage will be run after this.

Once the boxes around the sections are where you want them, click the "Recalculate" button. The normalization parameters will be recalculated and the rough-normalized result will be displayed in a sagittal view. If the rough normalization is adequate, then select only the "Fine-Tuning" check box (uncheck "Rough Norm.") and click "Calculate Normalization". The fine-tuning will then be applied and the result will be displayed.

Input parameters for normalizing brain regions

"Save name base" indicates the name that will be given to the normalization data file, plus a suffix indicating the series number.

"Brain template" is the name of the template to use for normalization. This must be a NIfTI format image in the "venv/braintemplates" folder.



ı iterative manner with successively finer resolution. The
ıe registration method in the "dipy" software package. To
ɔ enter 3 values for each parameter.

s to the maximum number of iterations for optimizing the
at each resolution level. For example: 10000, 1000, 10

rs to the spatial smoothing to apply to the image data at each

The parameter "divide per level" refers to how the data are sub-sampled for each level of resolution. For example, 4, 2, 1 indicates sampling every 4th point, then every 2nd point, and then every point.

As a result, the normalization process successively includes finer and finer details for matching the image data to the template.

IF **norm_bridge_ref** is defined in the database file:
A reference file to guide the normalization can be indicated in the database under the heading norm_bridge_ref. The reference series is indicated by its database number. That is, the reference is itself listed in the database in a different line.
If this heading exists in the database, and a reference file is not used, the value must be set to -1 (or any negative number).

Without a reference:
The normalization process involves successive steps of:
1) initial rough normalization of fMRI data to the template based on center of mass

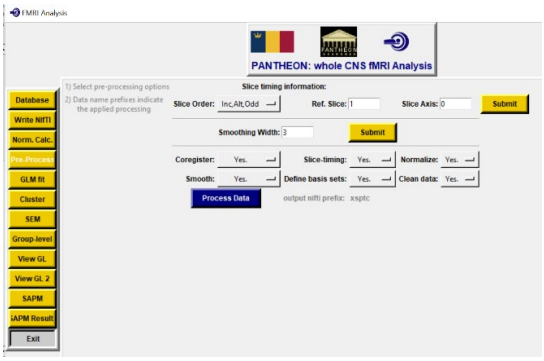
2) refining the match with a 3D affine transformation
3) refining again with a 3D rigid transformation
4) refining again with a 3D affine transformation

With a reference:
The normalization process involves successive steps of:
1) initial rough normalization of fMRI data to the reference scan based on center of mass
2) refining the match with a 3D affine transformation
3) refining again with a 3D rigid transformation
4) refining again with a 3D affine transformation

5) initial rough normalization of reference scan to the template based on center of mass
6) refining the match with a 3D affine transformation
7) refining again with a 3D rigid transformation
8) refining again with a 3D affine transformation

9) calculate the combined transformation from the fMRI data to the template

7. Pre-processing your data – the "Pre-Process" page



Pre-processing of fMRI data has many options, depending on your data, the type of study, etc. This page enables you to choose the pre-processing steps to apply, and choose options for how they are applied.

The top row of entry boxes and buttons is for entering information used in the slice-timing correction.

It is necessary to correctly indicate the order in which the slices were acquired, for the slice-timing correction to be done. The "Ref. slice" refers to the slice in each volume that you will use as a reference point in time, and the other slices will be interpolated to the times that the reference slice was acquired. Since data can be acquired in any orientation, but they are put into RAS orientation during the conversion to NIfTI format, it is necessary to indicate which dimension is the slice dimension (x:0, y:1, or z:2)

The second row is for indicating the smoothing width (if it is applied) in terms of the full-width at half-maximum of a gaussian smoothing kernel, in the units of voxels in the data.

The lower section is for indicating if you want to apply co-registration (motion correction), slice-timing correction, normalization (applying the normalization that was already calculated previously), spatial smoothing, defining the basis sets (for GLM analysis and removing physiological noise effects), and Cleaning the data (removing the effects of physiological noise).

Each option has the choice of "yes" – apply this step, "no" – do not apply this step, and "done" – this step has already been done, don't do it again, but use the data that already has this step applied.

The results of each step are written out as NIfTI format images, with the name prefixed with a letter according to each step. The resulting data prefix is indicated in the last section on this page, next to the "Process Data" button.

When you have made your choices, hit the "Process Data" button and the selected steps will be applied to all of the database entries that you indicated on the Database page.

Information about basis sets

The basis sets are defined based on the paradigm sheet in the database excel file. The timing described in that sheet is convolved with the canonical hemodynamic response function (HRF), and the result is added to the sheet in the database file. The basis sets also include models of bulk motion, as determined from the co-registration step, and also physiological noise based on common features of time-courses in white matter regions. In spinal cord and brainstem regions, specific regions of white matter are selected to identify potential global noise effects. For brain regions a white matter map is used to extract time-series responses in white matter, and the first three principal components of the time- series responses are used as models of global noise.

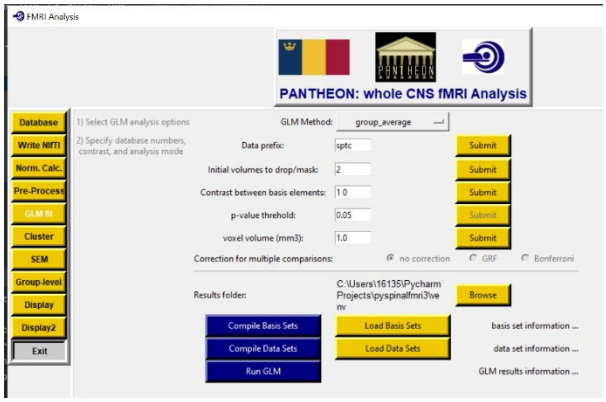
Each column in the paradigm sheet, other than the "dt" column, defines a paradigm. The way that multiple paradigms are used, such as whether they are combined or contrasted etc., is defined in the GLM section below.

The brain image template and white matter region definitions are in the folder "venv/braintemplates" in the pantheon package. These templates can be substituted for other templates, and the brain template can be selected in the normalization step. However, the white matter mask needs to be named avg152wm.nii" (for simplicity in the later step to generate models of the noise in the data). These files are from SPM12. If you do replace the white matter mask, keep the original in case you need to revert back to it. The anatomical region maps are taken primarily from the CONN15e software package as

described above in the section on normalization. The spinal cord and brainstem templates are in "venv/templates" in the pantheon package. These templates are a compilation of multiple sources including the PAM50 template and data from the Stroman Lab and multiple sources for defining anatomical regions, also as described earlier.

8. Fitting predicted BOLD responses to voxel data with a general linear model (GLM)

Voxel time-series responses can be fit to the model paradigms to identify which voxels have significant BOLD responses. This can be done in multiple ways, such as fitting the data from one person at a time, for multiple fMRI runs at a time, or with data combined across a group.



The "GLM method" is selected from a pull-down menu. The options are:
"group concatenate by person average" - average repeated runs in each person, but then concatenate the data from individual participants across the group to create one long time series
"group average" – average all of the runs for all participants to create one time-series response for the group
"concatenate by person" – concatenate repeated runs in each person, and analyze the data for each person separately
"concatenate group" – concatenate all runs in all participants to create one long time series "average by person" – average repeated runs in each person, and analyze the data for each person separately

"Data prefix" corresponds with the pre-processing steps that were applied to the data that you want to analyze, and is used to identify the NIfTI file names to read.

"Initial volumes to drop/mask" – the number of volumes at the start of the time-series acquisition that are expected to have less T1-weighting than later volumes, and are in a non-steady-state. Typically 2 volumes is sufficient. This number of volumes at the start of the time series are replaced with a copy of the first steady-state volume. That is, with the value set at 2, the first two volumes are replaced with copies of volume 3. This is to avoid non-steady-state volumes while also avoiding the problem of needing to shift the timing of stimulation/task periods that occurs when the initial volumes are removed instead of being replaced.

"Contrast between basis elements" – The significance of the fit of the time-series BOLD response in each voxel to the model paradigms can be assessed many different ways. The "contrast" between the paradigms that are defined in the database file is a set of numbers, one for each paradigm, to indicate how the paradigms are used. For example, if you simply want to know if the magnitude of the fit (the β value) to the paradigm is significantly different than zero, then the "contrast" would be just 1. If you have two paradigms, the data will be fit to both paradigms, but you can look at the significance of the fit to one of the paradigms, to determine if the magnitude is significantly different than zero. If the paradigm you are interested in is the first one, then the contrast would be 1 0, and if it is the second paradigm that you are interested in, it would be 0 1. You can run the analysis with different contrasts and save the results with different names. You could also ask whether the combined fit to both paradigms is significantly different than zero, or if the difference between the magnitudes of the fits between the two paradigms is significantly different than zero, and then the contrasts would be 1 1 for the combination or 1 -1 for the difference. Enter one value for each paradigm that you have defined. The contrasts are not applied to the terms that are added to the basis set, such as models of noise and

the constant offset value.

"p-value threshold" - This value specifies the statistical significance threshold for inferring that the result in each voxel is significant. For a single comparison this term is often set at 0.05, but for multiple comparisons with thousands of voxels the value needs to be set lower. However, the value entered here refers to the corrected p-value, after methods to correct for multiple comparisons are applied, as described below.

"voxel volume (mm3)" - this is an estimate of the voxel volume to be used with the Bonferroni method of correcting for multiple comparisons, as described below. For brain data the normalized volume is 8 mm3 (2 mm x 2 mm x 2 mm voxels) and for brainstem and spinal cord data the normalized volume is 1 mm3 (1 mm x 1 mm x 1 mm voxels). However, this is not the actual voxel volume in the original data.

"Correction for multiple comparisons" - select one of the radio buttons to indicate how you want to correct for multiple comparisons.
The choices are:
"none" - do not correct for multiple comparisons, just use the indicated p-value threshold
"GRF" – use gaussian random field (GRF) theory to estimate the smoothness of the data and estimate the actual number of independent statistical comparisons being made in the analysis of the data set. The actual p-value threshold is adjusted from the indicated value to correct for multiple comparisons based on this estimate.
"Bonferroni" – estimate the number of statistical comparisons as the total volume of data, divided by the voxel volume entered above. The actual p-value threshold is adjusted from the indicated value to correct for multiple comparisons based on this estimate.

"Results folder" - select where you want the results to be saved

"Compile Basis Sets" – this step is to collect together the definitions of the main effects basis sets (as determined from the paradigms defined in the database file) and the models of motion and noise. You will be prompted to specify a name and location for saving this file.

"Load Basis Sets" – if this basis sets have already been compiled you can load them without compiling them again. You will be prompted to select an existing basis set file.

"Compile Data Sets" – this step collects the data from all of the clusters in all of the data sets. You will be prompted to specify a name and location for saving this file.

"Load Data Sets" – if this data sets have already been compiled and saved you can load them without compiling them again. You will be prompted to select an existing data set file.

"Run GLM" – run the GLM fit of the specified basis sets to the selected data, using the parameters that you have specified.

9. Defining regions of interest and sub-regions by means of clustering – the "Cluster" page

Not all analysis methods are practical to apply on a voxel-by-voxel level, as with the GLM fitting described in the previous section. For connectivity analyses and investigations of BOLD time-series response characteristics, it is useful to identify regions of interest based on the anatomy. It is not expected that entire anatomical regions will have the same functions or the same connectivity with other regions, it is well known that some regions have somatotopic subdivisions, or functional subdivisions. The "Cluster" page is for extracting time series data from voxels within predefined anatomical regions, and applying k-means clustering to identify subregions based on function.

To apply clustering, it is first necessary to create a "network model" in the form of an Excel file. The Excel file needs to contain two sheets, named "connections" and "nclusters". In the **connection** sheet the network is defined by listing each "target" region (the region receiving input) and each corresponding "source" region (the regions providing input). The columns must include a number column, then a column with the heading "target", followed by columns with headings "source1", "source2", etc. An example is shown below. Regions can be both targets, and sources for other regions. It is important to note that the region names used in the network model file must exactly match the region names used in the region definition file "…/venv/templates/wholeCNS_region_definitions_cordsegments.xlsx". The names are case sensitive.

Table: An example of a network definition in the "connections" sheet

	target	source1	source2	source3	source4
0	C6RD	NRM	NGC	DRt	
1	DRt	C6RD	PAG	LC	
2	Hypothalamus	NTS	PAG	LC	
3	LC	Hypothalamus	NTS	PBN	
4	NGC	C6RD	PAG	NTS	NRM
5	NRM	C6RD	PAG	NTS	NGC
6	NTS	PAG	PBN		
7	PAG	Hypothalamus	Thalamus	NTS	
8	PBN	NTS	PAG	LC	
9	Thalamus	C6RD	LC		

The **nclusters** sheet describes the number of subregions for each region included in the network model. This sheet must contain a numbers column, and "name" and "nclusters" columns. The "name" column contains the name of each region in the network model in the connections sheet, and the "nclusters" column contains the number of subregions to define for each region. Again, the names must exactly match the names used in the connections sheet, and in the regions definition file.

Table: An example of an "nclusters" sheet

	name	nclusters
0	C6RD	5
1	DRt	5
2	Hypothalamus	5
3	LC	5
4	NGC	5
5	NRM	5
6	NTS	5
7	PAG	5
8	PBN	5
9	Thalamus	5

The name of the network model file will be entered in the "Cluster" page as the "Network Model" parameter, by clicking the "Browse" button and selecting the file.

"Data name prefix" corresponds with the pre-processing steps that were applied to the data that you want to analyze, and is used to identify the NIfTI file names to read.

"Cluster definition name" - this is the name of the file that will contain the cluster information for the network model you have selected. You can enter the name manually (and hit the "Submit" button), or use the "Browse" button to specify a name. This can be an existing file if you are using a previously defined set of clusters, or it can be a new file if you are defining the clusters.

"Region/cluster data name" - this is the name of the file that will contain the time-series data for the regions/clusters in the network model you have selected. The data will be organized by clusters, and will be concatenated across all participants in the specified database numbers. The data will be extracted and analyzed on a per-person basis for connectivity analyses. You can enter the name manually (and hit the "Submit" button), or use the "Browse" button to specify a name.

9. Connectivity analyses including structural equation modeling – the "SEM" page

It is not practical to carry out most forms of connectivity analyses on a voxel-by-voxel basis so the clusters defined in the previous section are used for connectivity analyses. The network to investigate is as described in the previous section.

Structural equation modeling is a family of statistical methods, but for the specific purposes of analyzing fMRI data it is used to explain the BOLD responses in each region of a network. The basic concept is that the BOLD responses reflect changes in metabolic demand of a region over time, and this relates to the total input signaling to each region. This input signaling comes from other regions in the network. We therefore model the BOLD responses as a linear combination of the BOLD responses in other regions, under the assumption that these BOLD responses also reflect the output signaling from each region. For example, if region A has a time-series response, S_A, and regions B and C are modelled as providing the input signaling, then $S_A = \beta_{BA}S_B + \beta_{CA}S_C + err$, where "err" represents the residual (unexplained) signal variation (23). The SEM method is to fit the data to the network for each "target" region, and determine the β values for each connection. We can think of this as identifying relationships between BOLD responses in different regions, even if we cannot specifically say that one region is providing the input signaling to another region. If there are relationships between BOLD responses in different regions we can infer that the regions are part of a coordinated network. The details and validation of the method have been published previously (8,23).

The method is set up in Pantheon to combine data from the same person across multiple runs (if there are multiple runs), and to calculate β values for each connection in the network for each person (i.e. each subject or participant). Data can also be extracted from each run within specific time periods or "epochs" to allow for the possibility that connectivity (β) values may change during different conditions within the stimulation paradigm or resting-state.

The significance of the β values can be assessed based on either the consistency, with the group average being significantly different than zero, or with β values being correlated with a covariate such as performance on a task, pain ratings, etc. However, the significance and display of the results is done on a later page.

Inputs:

On the "SEM" page the first input is the network model definition. This is entered into the Network Model box, and is an excel file as described for the "Cluster" page.

Next, the "Data prefix" corresponds with the pre-processing steps that were applied to the data that you want to analyze, and is used to identify the NIfTI file names to read.

"Cluster definition name" - this is the name of the file that contains the cluster information for the network model you have selected. You can enter the name manually (and hit the "Submit" button), or

use the "Browse" button to specify a name. This file was defined previously in the "Cluster" step.

"Region/cluster data name" - this is the name of the file that contains the time-series data for the regions/clusters in the network model, and that was defined in the "Cluster" step previously. You can enter the name manually (and hit the "Submit" button), or use the "Browse" button to specify a name.

"Epoch center times" can be a single number, or a list of numbers separated by commas. Each number specifies the center volume of a time period of length "Epoch length" (described below). This is to enable data to be analyzed only from specific time periods (such as during a task, between tasks, etc.). A set of results will be generated for each time that is entered. If the entire time period spanned by the paradigm is to be used, then the Epoch center should be set to the middle volume of the time series.

"Epoch length" is a single number that specifies the duration of each "epoch" to be analyzed. If the entire time period spanned by the paradigm is to be used, then the Epoch length should equal the number of volumes in the entire time series.

"Results save folder" - this is the name of the directory where the results will be saved. The location can be changed by selecting the "Browse" button and selecting a folder from a dialog box that will open.

"tag for results names" is simply a string of text that the user would like included in the name of results files. This is to ensure that the results files can be identified and distinguished from other results files. This name should not include any punctuation or spaces, and should not be excessively long. For example, if the data being analyzed are from the healthy control (HC) group and the stimulus that was applied is called "stim1", this text might be "HC_stim1". The text can be entered into the box, and then the user must press the "Submit" button.

Analysis Options:

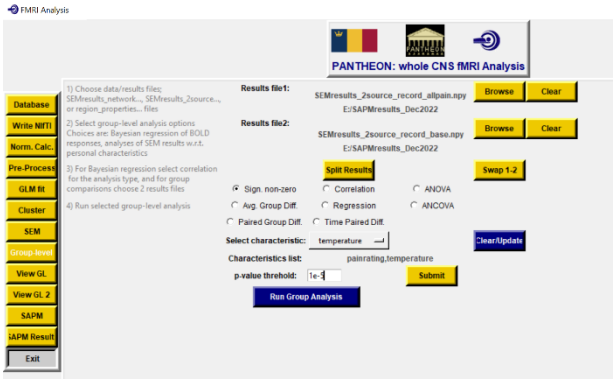
Var/Cov : This button runs the variance and covariance analysis, comparing all combinations of regions/clusters in the data set specified by "Region/cluster data name".

2-source SEM: This button runs an SEM analysis with every possible combination of two different clusters as sources, for every cluster as the target. The network is not used for this analysis, except that the regions listed in the network model are the regions that are tested. For a total of N clusters (where N is the number of regions times the number of clusters per region), this analysis will test Ntotal combinations, given by Ntotal = N (N-1) (N-2)/2 . For example, for a network consisting of 10 regions, and each region is divided into 5 clusters, then N = 50 and Ntotal = 58,800 combinations. This would be repeated for each "epoch" indicated by the "Epoch center times". The calculations are also carried out separately for each participant in the data set.

Network SEM : This button runs the full SEM analysis on the network specified in "Network Model". The analyses are carried out for one target region at a time, using all possible combinations of clusters for each region. The analyses are also carried out separately for each participant in the data set. Since this analysis can take some time (a few hours, depending on the data set), intermediate results are saved at each stage, when possible, and the analysis can be resumed from the last completed stage if the run is interrupted. This is done by checking the "Resume previous" box before pressing the Network SEM button.

10. Compiling the results – the "Group-level" page

In order to understand the results it is typically necessary to look at overall group trends and patterns, and to compare groups or conditions. The Group-level page includes methods for doing group-level analyses, or group-level comparisons for results that were generated on the GLM fit or SEM pages.



The analyses are based on the fit parameters for each person in the data set, and can be assessed for significance based on the difference from zero, or the differences between two groups (paired or unpaired), or based on correlations with covariates.

Results file 1: This input specifies the results to analyze for the first group (or the only group). The file name can be selected by pressing the "Browse" button. The value can be cleared, so that no input is given by pressing the "Clear" button. When the analysis that is selected (below) is not a group comparison, only Results file 1 is used.

Results file 2: This input specifies the results to analyze for the second group when group comparisons are to be done. The file name can be selected by pressing the "Browse" button. The value can be cleared, so that no input is given by pressing the "Clear" button.

The files selected in Results file 1 and Results file 2 can be switched by pressing the "Swap 1-2" button. The purpose is to reverse the comparison for group comparisons, or to change which file is used for single-group analyses.

The "Split Results" button will take two study groups from within one results file (for example all data sets were analyzed together as one group) and will split the results into two groups. The characteristic that is used to divide the groups is specified with the "select characteristic" option described below. In order to divide the data set based on this characteristic, the data sets must have two distinct values of this characteristic. For example, if "studygroup" was selected as the characteristic, the data could only be divided into two groups if there are only two distinct values for "studygroup" identified in the data. Continuous values (such as age, weight, pain rating, test scores etc.) cannot be used to divide groups into two clear sets. However, a characteristic could be added to the database file to label the data sets based on some feature, such as age above or below some value, etc.

Analysis type

The type of analysis to be carried out is identified with radio buttons:

"Sign. non-zero" indicates a single-group analysis based on the average fit parameters being significantly different than zero.

"Avg. Group Diff". is a group comparison based on the average of group 1 (Results file 1) minus the average of group 2 (Results file 2)

"Paired Group Diff". is a group comparison based on the difference between each value in group 1 minus the value in group 2. This requires paired data, and assumes that the order of the participants is identical in the two data sets.

"Correlation" indicates a single-group analysis based on the correlation between fit parameters and characteristic values across participants in the data set.

"Regression" indicates a single-group analysis based on the linear regression between fit parameters and characteristic values across participants in the data set.
"Time Paired Diff." is an analysis of time-series BOLD responses based on the difference between the BOLD signal intensity in time points between the two groups.
ANOVA: indicates an analysis of variance. This requires two groups to be compared (used as one categorical condition) and a choice of categorical characteristic to be selected (such as sex, study condition, etc.)
ANCOVA: indicates an analysis of covariance. This requires two groups to be compared (used as a categorical condition) and a choice of continuous characteristic to be selected (such as age, performance on a task, etc.)

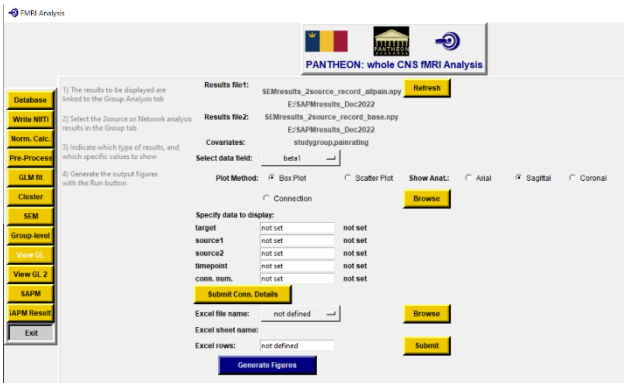
Select characteristic: indicates the pull-down menu which lists the choices of characteristics that have been loaded from the database file. Any characteristics that have been selected are listed next to the "Characteristics list" title. Multiple characteristics can be selected.

"p-value threshold" indicates the p-value to use for inferring significance, for the purposes of writing out results. This value can be entered manually, and then the "Submit" button must be pressed.

Once all the desired parameters have been selected, the analysis can be run by pressing the "Run Group Analysis" button.

The results are written to .npy files in the folder containing Results file 1.

11. Viewing the results – the "View GL" page



The results generated in the Group-level page are likely to be complex and to contain a tremendous amount of information, and these results need to be displayed in some comprehensible way so that they can be understood, and described to others (papers, presentations, etc.).

The View GL page (View Group Level) provides methods for loading and displaying the group level results. This page is very closely linked to the Group-level page. You can copy the definitions of file names etc. from the Group-level page over to the "View GL" page by pressing the "Refresh" button in the upper right part of the page. This fills in the values for "Results file 1" and "Results file 2", which have the same definitions as on the Group-level page. These names refer to the files containing the group-level results to view.

"Covariates" lists the characteristic values that were selected on the Group-level page. These covariates will again be used for plotting values etc.

"Select data field": identifies the pull-down list that has been populated with the names of the results values in Results file 1. This list will depend on the type of analysis that was done (GLM, SEM, etc.).

"Plot Method": gives options for showing a "Box Plot" or a "Scatter Plot" or a "Connection" plot, when the results are plotted.

The "Browse" button next to the "Connection" plot option is for selecting an Excel file for defining how to display the connection plot. This Excel file must be defined to correspond with the network definition file. It contains columns with the headings "name", "posx", "posy", "labeloffset_x", "labeloffset_y", and "outputangle". There must be a row for each anatomical region included in the network file, including the latent inputs (which are referred to as "intrinsic" and labeled int0, int1, int2 etc.).
Each row then contains the name of the anatomical region (exactly matching how it is specified in the network file), then the relative position of the region in the plot, with the values "posx" and "posy". These are values between 0 and 1, and refer to the relative left-right and bottom-top position in the plot.
The position of the region label, relative to the position of the region, is indicated with "labeloffset_x" and "labeloffset_y". Finally, the value of "outputangle" refers to where on the periphery of the region the output signaling is drawn as exiting the region. This value is in degrees.

Network plotting definition example (Excel file):

name	posx	posy	labeloffset_x	labeloffset_y	outputangle
C6RD	0.65	0.15	0	-0.05	90
DRt	0.35	0.25	0	-0.05	0
Hypothalamus	0.3	0.85	0	0.05	0
LC	0.8	0.65	0.05	-0.05	180
NGC	0.7	0.4	0	-0.05	-90
NRM	0.45	0.4	0.05	0	-90
NTS	0.1	0.6	-0.05	-0.05	0
PAG	0.6	0.75	0.025	-0.05	-90
PBN	0.25	0.5	0	-0.05	0
Thalamus	0.6	0.95	0.05	-0.05	-90
int0	0.8	0.15	0	-0.05	180
int1	0.8	0.55	0	-0.05	90
int2	0.1	0.7	0.05	-0.05	-90

"Show Anat." gives options for showing anatomical regions in an "Axial", "Sagittal", or "Coronal" slice.

Under the heading of "Specify data to display" there are multiple ways of specifying the results to show, and these can depend on the type of data being shown.
Values referring to the "target", "source1", "source2", "timepoint", and "conn. num." (i.e. connection number) can be specified manually. Multiple values separated by commas can be entered for each value. These values are exactly as listed in the Excel file containing the results of the Group-level analysis.
The Excel file that contains these results must be selected next to "Excel file name" by pressing the "Browse" button and selecting the file with the dialog box that opens.
The sheet name within this file can then be selected from the pull-down menu, and the sheet name will be listed next to "Excel sheet name".

The results to show can also be specified by entering one or more line numbers, separated by commas, next to the label "Excel rows" and hitting the "Submit" button. The row numbers correspond to the Excel file. When a line number, or multiple numbers, is/are entered and the "Submit" button is pressed, then the values of "target", "source1", etc. above will be populated.

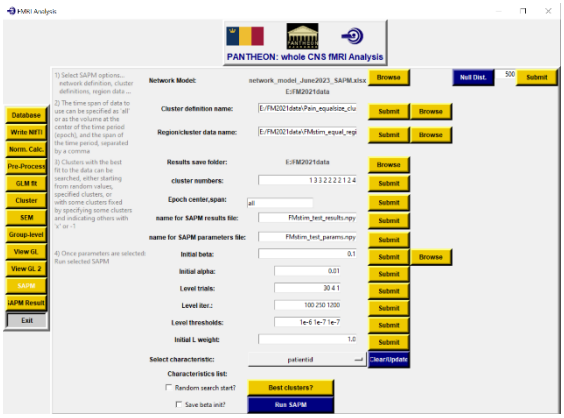
The results are then displayed when the "Generate Figures" button is pressed. The values that are used are those that are listed in the "target", "source1", etc, fields when "Generate Figures" is pressed.

The results figures are shown on the "View GL 2" page.

12. Viewing the results – the “View GL 2” page

The “View GL 2” page contains only two figure windows. Plots are displayed in the left window and anatomical slices are displayed in the right window. Copies of the figures can be saved in “.svg” format by pressing the “Save Plot” or “Save Image” buttons below each window.

13. Structural and Physiological Modeling – the "SAPM" page



SAPM is a connectivity analysis method that extends the SEM method (8). SAPM combines *a priori* knowledge of anatomy, neurophysiology, and the relationships between physiological processes and blood oxygenation-level dependent (BOLD) MRI signal variations. This information is used to model the neural signaling underlying observed BOLD responses across interconnected networks of regions. The basic concept underlying SAPM is that since the BOLD response is known to relate to pre-synaptic input signaling and the metabolic demand that is driven by incoming neural signaling, the input signaling to each region, S_{input}, can be modeled as the sum of the outputs from other regions in the network:

$$S_{input} = M_{input}\, S_{output}$$

Here M_{input} is a matrix of "D" values which are the weightings of each incoming signal as they are summed to produce the net total input signaling from each region, and S_{output} is the output signaling from each region. Note that all D values are positive values and allow for the output signaling from a single region to contribute different amounts of input to other regions. The neural signaling here is represented as the relative drive in metabolic demand. That is, if the incoming signal gets larger (i.e. increased release of neurotransmitters at synapses on dendrites) it will produce an increase in the metabolic demand of the region. The output signaling from each region is modeled similarly as (Figure 5):

$$S_{output} = M_{output}\, S_{output}$$

Here M_{output} is a matrix of weighting factors which are related to how each input signal influences the output signaling from the region. A positive weighting factor corresponds with excitatory input (more input produces more output) whereas a negative value corresponds with inhibitory input (more input produces less output). The weighting factors in M_{output} are termed "DB" values because they are the product of D values mentioned above, and the B values which reflect how the incoming signal is converted to contribute to the output signal within each region.

'

Applying the SAPM method consists of defining the network model, defining the BOLD signal time-courses in each region, and then computing the D and DB values and also computing the latent inputs.

Inputs

Although the "Null Dist." option is near the top of the page, on the right side, this option is described at the end, under the heading "Running SAPM"

The "Network Model" input is the name of an Excel file that defines the network, as described above for the Cluster, and SEM pages. The network model file can be selected by pressing the "Browse" button.

"Cluster definition name" - this is the name of the file that contains the cluster information for the network model you have selected, as in previous pages. You can enter the name manually (and hit the "Submit" button), or use the "Browse" button to specify a name. This file was defined previously in the "Cluster" step.

"Region/cluster data name" - this is the name of the file that contains the time-series data for the

regions/clusters in the network model, as in previous pages. You can enter the name manually (and hit the "Submit" button), or use the "Browse" button to specify a name.

"Results save folder" - this is the name of the directory where the results will be saved. The location can be changed by selecting the "Browse" button and selecting a folder from a dialog box that will open.

"cluster numbers" is a list of numbers, one for each region in the network model definition. These numbers refer to the cluster number for each region. That is, these values specify the sub-region within each region. For example, if each region was divided into 5 clusters, each number is a value between 0 and 4. The SAPM values will only be computed for the network composed of these regions/sub-regions.
This leads to the question of how are the cluster numbers determined? This is described below under the heading of "Best clusters?"

"Epoch center, span" is a set of two values separated by a comma, and are used to specify the portion of the time series data to use. This is the same idea as the "Epoch center times" and "Epoch length" described on the SEM page. The first number is the volume number of the epoch center, and the second number is the span of the data as a number of volumes. However, if all of the data is to be used (i.e. no selected time period) then the word "all" can be entered instead. For example, if there are 100 volumes in a time series, one might want to analyze the first half of the runs and enter the values 25, 50 for the center and span of the data to use. Alternatively, if all 100 volumes should be used, then the word "all" could be entered instead of two numbers. It is necessary to press the "Submit" button after the values are entered.

"name for the SAPM results file": enter a file name ending with the extension ".npy" for saving the results. Do not include the directory name, only the file name is entered. The results will be saved in the "Results save folder" defined above.

"name for the SAPM parameters file": enter a file name ending with the extension ".npy" for saving the analysis parameters. Again, do not include the directory name, only the file name is entered. It is generally a good idea to use similar names for the "results" and "parameters" files because they are always used together in later steps. For example, if the results file was named "SAPM_study1_results.npy" then the parameters file could be named "SAPM_study1_params.npy". This is just for convenience later.

"Initial beta": The DB values are estimated using a gradient descent approach, starting from a random set of values. The value of Initial beta is used to scale the range of this initial guess. The randomly generated numbers have a normal distribution, with an average of zero and a standard deviation of 1.0. These random numbers are multiplied by "Initial beta" to make the range larger, or smaller. If Initial beta was set to zero, all of the starting guesses for DB values would equal zero. Large values can cause the gradient descent method to diverge and fail. Typically, the best number to use is about 0.1. A value that is considered large would be about 0.3, and a small value would be 0.01 (or zero). As described below, this procedure is repeated multiple times in order to find the (hopefully) overall best fit starting from different initial guesses.

"Initial alpha": This parameter is also used in the gradient descent method, and is a scaling factor that affects how much the DB values are changed on each step of the gradient descent. Larger values can produce a faster approach to the optimal value, but smaller values can give better accuracy. If the change in DB values on any step does not result in a better fit (i.e. the error increases) then the change in DB values is too large and the alpha value is decreased automatically, and the step is repeated. If multiple steps in a row produce improvements (i.e. smaller error) then the alpha value is automatically increased slightly to speed

up the fitting procedure. Typically, starting alpha values are about 0.1 or 0.01.

"Level trials": The gradient descent method can depend on the initial guess of starting DB values, so the procedure is repeated with multiple starting values. Three different stages of the procedure are used. The first stages starts with random DB values. The second stage starts from the best results of stage one. The final (third) stage starts with the best result of stage two. The number of "level trials" refers to how many sets of starting values are used in each stage. These are entered as a list of three integers, separated by commas or spaces. Typical values are such as 30, 4, 1. Larger numbers could provide more accuracy, but will take longer.

"Level iter.": For each of the sets of starting DB values for each trial, the gradient descent procedure is repeated up to a maximum number of iterations, or until the error term is not decreasing a sufficient amount on each step (for how much is a "sufficient" amount, see the "Level thresholds" below). These are entered as a list of three integers, separated by commas or spaces. The initial stages can have fewer iterations for the process of searching for which sets of starting DB values provide converging results. Later stages, starting from the most promising sets of DB values, are done with larger numbers of iterations. The final stage has the highest number of iterations. Typical values are such as 100, 250, 1200. Larger numbers can provide better accuracy, but will take longer.

"Level thresholds": This parameter is also used in the gradient descent method, and refers to how large the change in error term needs to be for the procedure to be continued. These are entered as a list of three floating-point numbers, separated by commas or spaces. If the error term is only changing by a very small amount on each iteration, then the DB values are not really improving, and there is not much point in continuing the procedure. However, sometimes the improvements are small for a number of iterations and then much larger reductions in the error are produced by each iteration. Eventually, no improvement in DB values can be found, if the optimal values (or a local minimum) have been found. Typical values are such as: 1e-5, 1e-6, 1e-7. Smaller values can produce better accuracy, but will take longer.

"Initial L weight" : this value is the scaling factor for L1 weighting for the gradient descent method. The gradient descent method is to make the error term as small as possible, but there is also a "cost" function that is minimized. This "cost" function is the average magnitude of the DB values. As a result, the method tries to find the combination of DB values that have a balance of producing the least error, while also having the smallest average magnitude. Smaller L weight values make the cost function have less of an influence on the outcome, and larger L weight values make this cost function have more of an influence. The best L weight value to use can vary with the size of the error term, and the magnitude of the DB values, but values around 1.0 appear to be effective.

"Select characteristic" : Similar to the characteristics that are selected in previous steps, any covariates that are to be used in the analysis can be selected from the pull-down menu. The selected terms are listed below next to the "Characteristics list" label.

Running SAPM

"Best clusters?" is a method for searching for the set of cluster numbers that provides the best fit of the BOLD time-series data to the network model. The starting choice of cluster numbers can be the values specified in "cluster numbers" or the "Random search start?" option can be selected so that the starting set of cluster numbers is randomized. This method uses the settings "Level trials", "Level iter", and "Level thresholds" that are described above, and these values can be adjusted to speed up the search procedure at the cost of less precision in the fit values. For example the search can be run with fewer trials during each

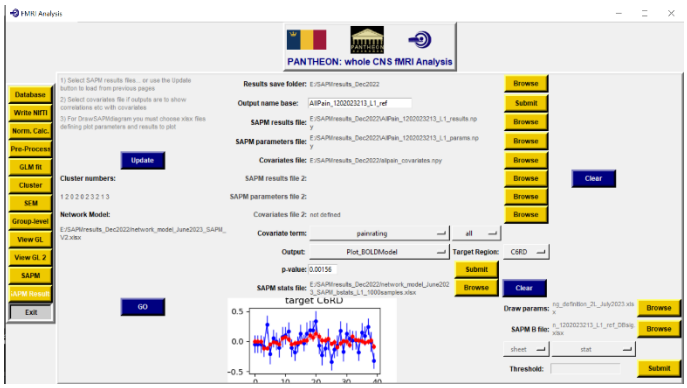
stage, and fewer iterations on each trial. The results can vary depending on the starting values. There is no guarantee that there is indeed a single "best" set of clusters, but the method does identify clusters that fit well to the model.

"Run SAPM" : This button starts the full SAPM analysis with the selected set of clusters. If desired, the initial guess of DB values can be saved in a file by selecting the "Save beta init?" option.

"Null Dist." This mode of running SAPM uses the specified Network Model but computes the range of DB values that could be obtained by random chance, using random data values. This is for the purpose of assessing the significance of the results that are obtained with SAPM using a real set of data. This method also uses the settings "Level trials", "Level iter", and "Level thresholds" that are described above. The number of null distributions to simulate should be a fairly large number, such as 1000, to obtain representative statistical distributions. The number of "null" simulations is entered in the box to the right of the "Null Dist." button, and then the "Submit" button must be pressed. The results are saved in the same directory that contains the Network Model file, with a file that that is the Network Model name with "_bstats" append plus the extension ".npy". A version is also saved in Excel file format, with the extension ".xlsx". If there is an existing Excel file with the same name, and there is a problem writing to it, the name is also appended with the date and time so that an error does not occur and the results are lost. This Excel version of the results might be needed when assessing the significance of the SAPM results.

14. Viewing SAPM results – the "SAPM Results" page

The "SAPM results" page is closely linked to the previous "SAPM" page, so the first step is to press the "Update" button in the upper left region of the page. This copies the values from the SAPM page to the SAPM results page, as much as possible. It is important to check that the "Cluster numbers" and "Network Model" listed below the "Update" button are correct for the analysis that you want to do. If they are not correct, go back to the "SAPM" page and change the values there.



In the middle of the page, at the top, the "Results save folder" is specified. Use the "Browse" button to change this folder name if wanted. As the name suggests, this is where the results will be saved.

Inputs

"Output name base" specifies a portion of the name that will be used for the output files. This should be some specific name to distinguish the results from other analysis. For example, it could include a name indicating the study type and group, and could list the cluster numbers, such as "Stim1_HC_0131232134". It can be whatever you want except it must not contain spaces or punctuation characters.

"SAPM results file" and "SAPM parameters file" are the same names as the results and parameters files that were selected on the SAPM page. Use the "Browse" buttons to change these by selecting other files.

"Covariates file" : The covariates file that was generated on the SAPM page can be specified here for later use the analysis. This is optional.

If desired, a second set of results can be specified, with "SAPM results file 2", "SAPM parameters file 2" and "Covariates file 2", for the purposes of comparing sets of results from different data sets.

"Covariate term" : A covariate term of interest (for analysis) can be selected from the pull-down menu, and this relates to the values that were loaded from the "Covariates files" specified above.

The "Output" pull-down menu can now be used to select a number of different types of results to display. Select the options on the pull-down menu from these choices: B_significance, B_regression, Plot_BOLDModel, Plot_SourceModel, DrawSAPMdiagram, DrawAnatomy_axial, DrawAnatomy_sagittal, Group_Diff, Paired_Diff, and Regress_diff_v_diff. Some of these options have other inputs that need to be defined, as described below, and the output is not generated until the "GO" button is pressed (lower left part of the page)

Within these options:

"B_significance" calculates the group average DB, D, and B values are their significance (based on the probability of being equal to zero) using a Student's T-test. The significant results are written to Excel files.

"B_regression" calculates the linear regression between the DB, D, and B values and the selected

covariate values across participants. The significant results are written to Excel files.

"Plot_BOLDModel" creates a plot of the measured and modeled BOLD responses for the selected region. The results are displayed in the window below and also written to an ".svg" file.

"Plot_SourceModel" creates a plot of the input and output signaling for the selected region, with the measured BOLD responses plotted as well for comparison. The results are displayed in the window below and also written to an ".svg" file.

"DrawSAPMdiagram" creates a figure showing the network model and the significant connections within it. The results are written to an ".svg" file. For this output it is necessary to specify the Excel file that defines how to plot the connectivity diagram, using the "Draw params" input in the lower right corner of the page (more details below). Use the "Browse" button to select the file. This Excel file is as described above for plotting the network regions on the "View GL" page.

"DrawAnatomy_axial" generates images of each cluster in "Cluster numbers". The images are gray-scale axial slices of the template image, with the cluster voxels marked in color. The slice shown is through the center-of-mass of the cluster. The images are written as ".svg" files.

"DrawAnatomy_sagittal" generates images of each cluster in "Cluster numbers". These are as described above for DrawAnatomy_axial, except now the slices are sagittal. The images are written as ".svg" files.

"Group_Diff" calculates the differences between the group average values for the two sets of data that are specified. Significance is calculated using a two-sample Student's T-test. The results are written to an Excel file.

"Paired_Diff" calculates the paired differences between values in the two sets of data that are specified. Significance is calculated using a Student's T-test based on the averages and standard errors of the paired differences between the two data sets. The results are written to an Excel file.

"Regress_diff_v_diff" calculates the linear regression between the difference in paired values in the two sets of data, and the difference in covariate values for the two sets of data. For example, if the covariates are some test performance score for two conditions, the difference between scores for each person can be indicated with ΔS. Comparing these values with the differences in DB values, ΔDB, between the two conditions for each person, the linear regression that is done would be ΔDB vs ΔS. The results are written to an Excel file.

"p-value" : This value indicates the p-value to use for the statistical comparisons. The value can be entered manually and then the "Submit" button must be pressed. The choice of p-value must be adjusted to account for multiple comparisons. That is, this is the uncorrected p-value.

"SAPM stats file" : This input refers to the Excel file that was generated with "null" tests to determine the connectivity values that could occur by random chance. This input is important for accounting for possible biases in values away from zero. That is, depending on the network model, it is possible that even with random data some connectivity values can tend to be slightly skewed to positive or negative values on average. Statistical comparisons to identify "significant" connectivity values should be compared against these values that are obtained with "null" tests. If no file name is specified here, the values are compared with zero for tests of significance. Detailed explanations are in "Structural and Physiological Modeling

(SAPM) for the analysis of functional MRI data applied to a study of human nociceptive processing", P W Stroman, M Umraw, B Keast, H Algitami, S Hassanpour, J Merletti (under review)

"Draw params" (lower right portion of the page) This input is mentioned above, and specifies the Excel file to define how networks should be displayed. This Excel file is as described above for plotting the network regions on the "View GL" page. Use the "Browse" button to select the file.

"SAPM B file" This input refers to the Excel file containing the results that are to be plotted as significant connections in the SAPM diagram. Use the "Browse" button to select the file.

"sheet" and "stat" pull-down menus below the "SAPM B file" name are used to specify specifically which data sheet, and which stat value in the Excel file to plot in the SAPM diagram.

"Threshold" indicates how to limit the values that are plotted. Typically, the only values in the Excel file are those that are significant. However, the user may choose to list more values by setting a higher p-value threshold, and then at this stage could select a higher T-value threshold for plotting. The input to "Threshold" is a text string indicating a range and whether the range applies to the magnitude ("mag") or the absolute value ("abs"). For example, the text could be as "abs > 3.0". If in this case the selected values to plot are T-values, then only results with absolute values greater than 3.0 would be shown on the plot. If all of the values are to be plotted, then "abs > 0" could be entered. If only positive values are to be plotted then the text entered could be "> 0" or "mag > 0". After the text is entered it is necessary to press the "Submit" button.

References

1. Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited--again. NeuroImage 1995;2(3):173-181.
2. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain connectivity 2012;2(3):125-141.
3. De Leener B, Fonov VS, Collins DL, Callot V, Stikov N, Cohen-Adad J. PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. Neuroimage 2018;165:170-179.
4. Bosma RL, Stroman PW. Assessment of data acquisition parameters, and analysis techniques for noise reduction in spinal cord fMRI data. Magn Reson Imaging 2014;32(5):473-481.
5. Powers JM, Ioachim G, Stroman PW. Ten Key Insights into the Use of Spinal Cord fMRI. Brain sciences 2018;8(9).
6. Stroman PW, Figley CR, Cahill CM. Spatial normalization, bulk motion correction and coregistration for functional magnetic resonance imaging of the human cervical spinal cord and brainstem. Magn Reson Imaging 2008;26(6):809-814.
7. Stroman PW, Kornelsen J, Lawrence J. An improved method for spinal functional MRI with large volume coverage of the spinal cord. J Magn Reson Imaging 2005;21(5):520-526.
8. Stroman PW, Powers JM, Ioachim G. Proof-of-concept of a novel structural equation modelling approach for the analysis of functional magnetic resonance imaging data applied to investigate individual differences in human pain responses. Hum Brain Mapp 2023;44(6):2523-2542.
9. Lang J, Bartram CT. [Fila radicularia of the ventral and dorsal radices of the human spinal cord]. Gegenbaurs Morphol Jahrb 1982;128(4):417-462.
10. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers, Inc.; 1988.
11. Millan MJ. Descending control of pain. Progress in neurobiology 2002;66(6):355-474.
12. Keren NI, Lozar CT, Harris KC, Morgan PS, Eckert MA. In vivo mapping of the human locus coeruleus. Neuroimage 2009;47(4):1261-1267.
13. Naidich TP DH, Delman BN, Sorensen AG, Kollias SS, Haacke EM. INTERNAL ARCHITECTURE OF THE BRAIN STEM WITH KEY AXIAL SECTIONS. Duvernoy's Atlas of the Human Brain Stem and Cerebellum. New York: Springer-Verlag/Wien; 2009. p p 79-82.
14. Leijnse JN, D'Herde K. Revisiting the segmental organization of the human spinal cord. J Anat 2016;229(3):384-393.
15. De Leener B, Levy S, Dupont SM, Fonov VS, Stikov N, Louis Collins D, Callot V, Cohen-Adad J. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 2017;145(Pt A):24-43.
16. Pauli WM, Nili AN, Tyszka JM. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. Scientific data 2018;5:180063.
17. Chiang MC, Bowen A, Schier LA, Tupone D, Uddin O, Heinricher MM. Parabrachial Complex: A Hub for Pain and Aversion. J Neurosci 2019;39(42):8225-8230.
18. Liebe T, Kaufmann J, Li M, Skalej M, Wagner G, Walter M. In vivo anatomical mapping of human locus coeruleus functional connectivity at 3 T MRI. Hum Brain Mapp 2020;41(8):2136-2151.
19. Stroman PW, Warren HJM, Ioachim G, Powers JM, McNeil K. A comparison of the effectiveness of functional MRI analysis methods for pain research: The new normal. PLoS One 2020;15(12):e0243723.
20. Gray's Anatomy: The Anatomical Basis of Medicine and Surgery. Williams PL, Bannister LH, Berry MM, Collins P, Dyson M, Dussek JE, Ferguson MWJ, editors. New York: Churchill-Livingstone; 1995. 975-1011 p.

21. Naidich TP, Duvernoy HM, Delman BN, Sorensen AG, Kollias SS, Haacke EM. INTERNAL ARCHITECTURE OF THE BRAIN STEM WITH KEY AXIAL SECTIONS. Duvernoy's Atlas of the Human Brain Stem and Cerebellum. New York: Springer-Verlag/Wien; 2009. p 79-82.
22. Lang J. Clinical anatomy of the cervical spine. New York: Thieme medical Publishers; 1993. 192 p.
23. Stroman PW. Validation of Structural Equation Modeling Methods for Functional MRI Data Acquired in the Human Brainstem and Spinal Cord. Crit Rev Biomed Eng 2016;44(4):227-241.