PANTHEON

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

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 tested somewhat on Mac environments. 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. That is, 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, and structural equation modeling (SEM)
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 (choose version 3.2.2)
nibabel
openpyxl
pandas
Pillow
pyinstaller
sklearn
xlrd (choose version 1.2.0)
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 - choose python3.6, or 3.8 also seems to work, from wherever this was installed with Anaconda

- 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:
    dicom2nifti
    numpy
    matplotlib (choose version 3.2.2)
    nibabel
    openpyxl
    pandas
    Pillow
    pyinstaller
    sklearn
    xlrd (choose version 1.2.0)
    dipy
    statsmodels

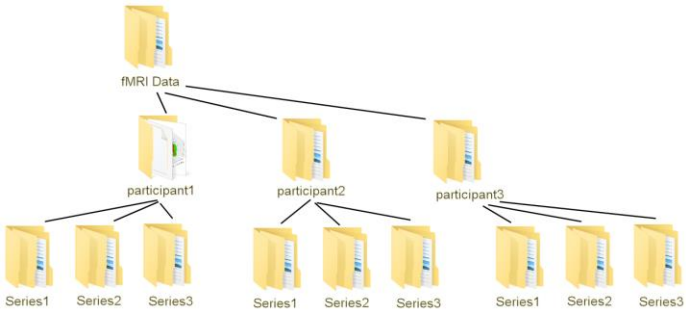
STEP 4
run pantheon.py in Pycharm

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 NIfTI file.



It is preferable, for ease of organizing your data, to keep all of the data folders for each participant 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 "database" file in Excel format. This is to make the 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.

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:

	patientid	studygroup	pname	seriesnumber	niftiname	TR	normdataname	normtemplatename	paradigms	datadir
0	pid1	sg1	e1	pnamseriesnum1	tbd	TR	tbd	ccbs	m1	paradigir	datad
1	pid2	sg2	e2	pnamseriesnum2	tbd	TR	tbd	ccbs	m1	paradigir	datad
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\participant1
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
normdatanam e	The name of the file that will contain the information needed to normalize your data. This	"TBD" or "undetermined" etc.

	will be filled in by Pantheon when the normalization parameters are computed.	
normtemplate name	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)

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.

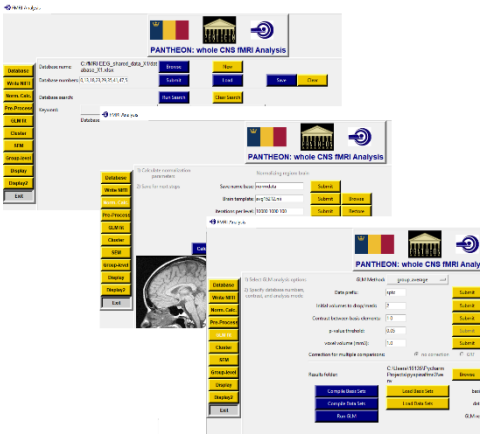
The example shown above describes 20 seconds when the metabolic demand is expected to be lower (such as a "rest" condition), followed by 20 seconds when the metabolic demand is expected to be higher (such as a stimulation condition), following by another period with lower metabolic demand.

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

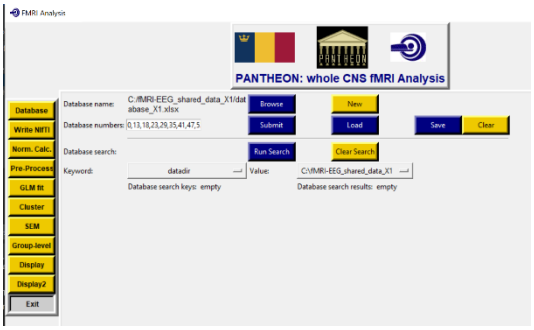


“Pages” refers to the different views that you can have in Pantheon. The buttons on the left side of each page take you to other pages, with other options and steps for preprocessing and analysis.

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.

5. 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.

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) (refs) and the CONN15e software package (refs), and the cord templates are a combination of the PAM50 template (ref Deleener) and templates created in the Stroman Lab (refs), and supplemented by other region maps (refs). 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.

---insert descriptions of templates from papers?----------

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, ref) software.

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 (refs).

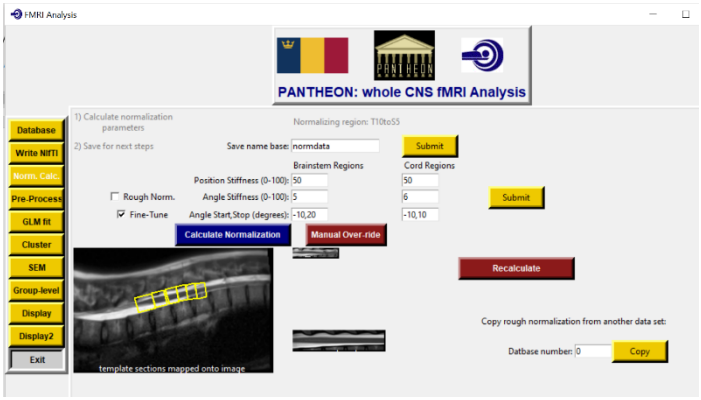
The page includes information messages such as near the top/middle the region to be normalized is indicated. If this does not match what you expect, make sure you have correctly entered your database name and database numbers on the first page.

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.



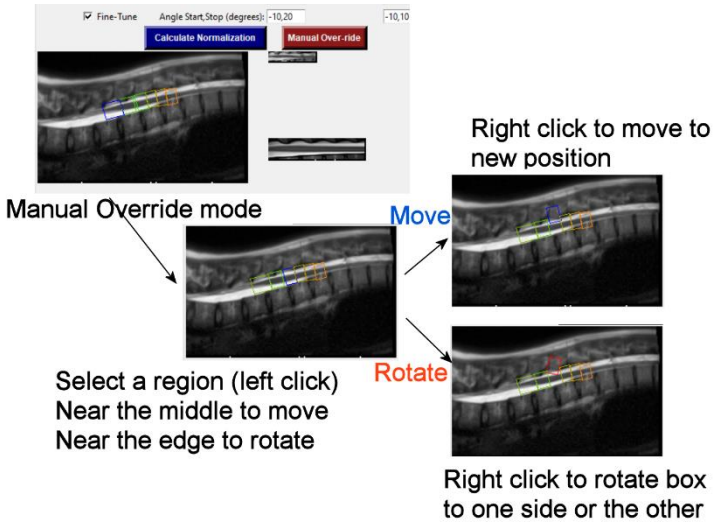
“Angle Stiffness”: this is similar to position stiffness as it refers to how much flexibility is allowed between the estimated angle for each section and the expected angle for each section.

“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.

The normalization process is applied in an iterative manner with successively finer resolution. The parameters are as required for the 3D affine registration method in the "dipy" software package. To indicate 3 levels of resolution, you need to enter 3 values for each parameter.

The parameter "iterations per level" refers to the maximum number of iterations for optimizing the match between the data and the template at each resolution level. For example: 10000, 1000, 10

The parameter "smoothing per level" refers to the spatial smoothing to apply to the image data at each level. For example: 3.0, 1.0, 0.0

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.

The normalization process involves successive steps of:
1) initial rough normalization 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

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

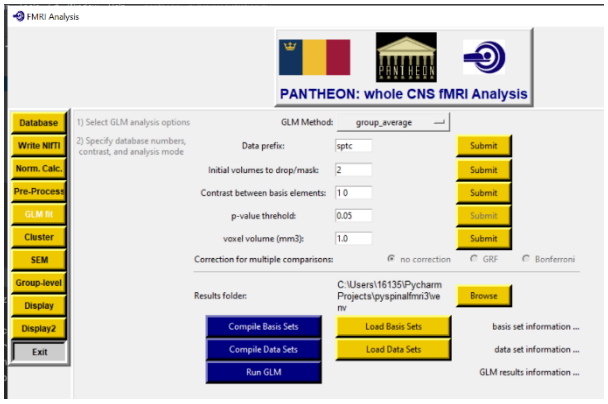
The basis sets are defined based on the paradigm sheet in the database excel file. The timing described in that sheet is convolved with the canonical hemodynamic response function (HRF), and the result is added to the sheet in the database file. The basis sets also include models of bulk motion, as determined 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 (ref).

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 (De Leener et al. ref), data from the Stroman Lab (refs) and multiple sources for defining anatomical regions (refs).

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.

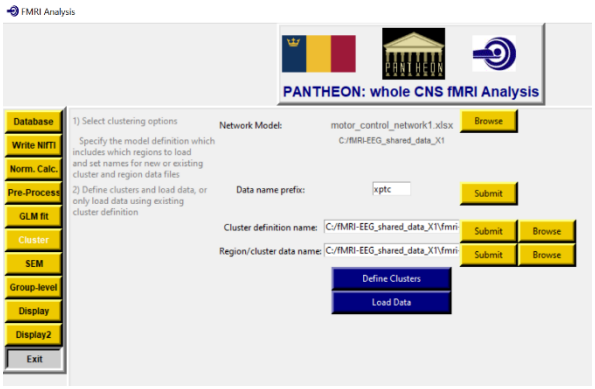
"Compile Data Sets" – this step collects the data from all of the clusters in all of the data sets. You will be prompted to specify a name and location for saving this file.

"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.

10. 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. On the “SEM” page the inputs are …