Development of a robust and reproducible preprocessing pipeline for Positron Emission Tomography (PET) data

Avneet Kaur

M.Sc. student,
Department of Computer Science,
University of Copenhagen

14th January 2022

Outline

- Introduction & Motivation
- Methodology
- Testing & Evaluation
- Results & Discussion
- Conclusion

Introduction

- PET is an essential neuroimaging tecnique used to quantify the concentration of molecular targets in the brain.

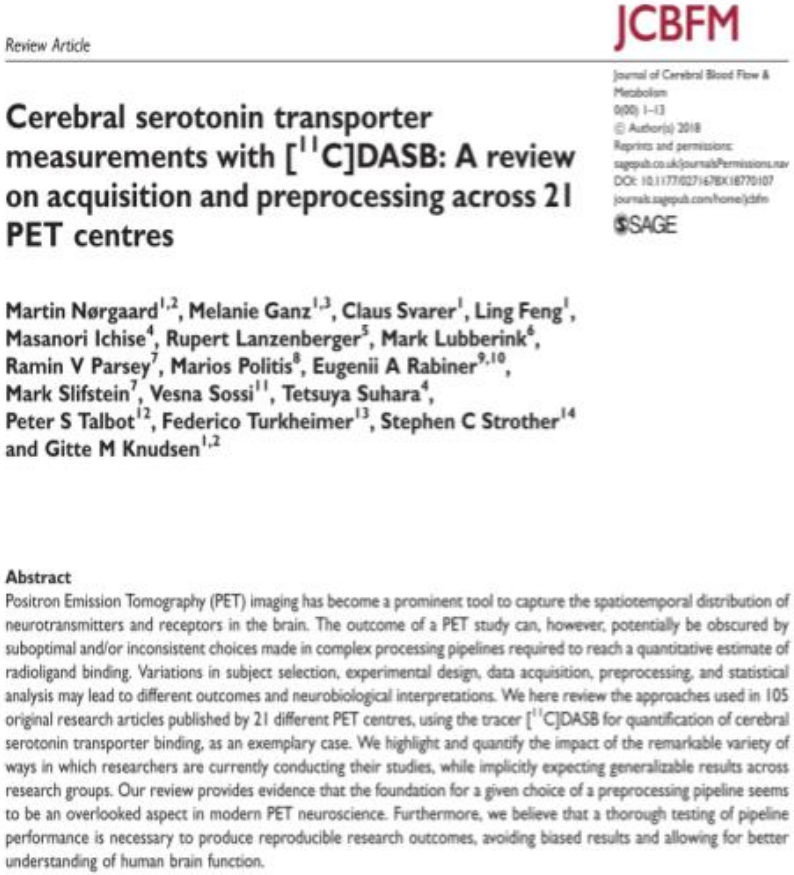
Figure 1: PET scanner

Introduction

- PET is an essential neuroimaging tecnique used to quantify the concentration of molecular targets in the brain.
- It has been used to study various disease and assess brain health.

Figure 1: PET scanner

Preprocessing strategies for PET data



Review Article

JCBFM

Journal of Cerebral Blood Flow & Metabolism
0(00) 1–13
© Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0271678X18770107
journals.sagepub.com/home/jcbfm

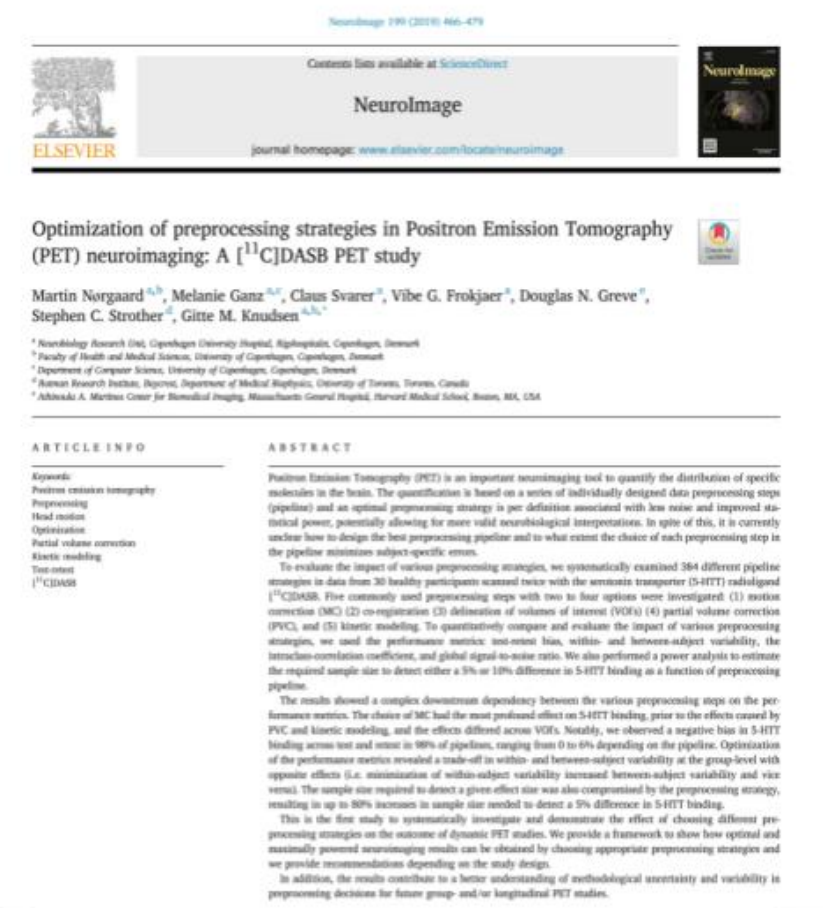
SAGE

Cerebral serotonin transporter measurements with [^{11}C]DASB: A review on acquisition and preprocessing across 21 PET centres

Martin Nørgaard[1,2], Melanie Ganz[1,3], Claus Svarer[1], Ling Feng[1], Masanori Ichise[4], Rupert Lanzenberger[5], Mark Lubberink[6], Ramin V Parsey[7], Marios Politis[8], Eugenii A Rabiner[9,10], Mark Slifstein[7], Vesna Sossi[11], Tetsuya Suhara[4], Peter S Talbot[12], Federico Turkheimer[13], Stephen C Strother[14] and Gitte M Knudsen[1,2]

Abstract

Positron Emission Tomography (PET) imaging has become a prominent tool to capture the spatiotemporal distribution of neurotransmitters and receptors in the brain. The outcome of a PET study can, however, potentially be obscured by suboptimal and/or inconsistent choices made in complex processing pipelines required to reach a quantitative estimate of radioligand binding. Variations in subject selection, experimental design, data acquisition, preprocessing, and statistical analysis may lead to different outcomes and neurobiological interpretations. We here review the approaches used in 105 original research articles published by 21 different PET centres, using the tracer [^{11}C]DASB for quantification of cerebral serotonin transporter binding, as an exemplary case. We highlight and quantify the impact of the remarkable variety of ways in which researchers are currently conducting their studies, while implicitly expecting generalizable results across research groups. Our review provides evidence that the foundation for a given choice of a preprocessing pipeline seems to be an overlooked aspect in modern PET neuroscience. Furthermore, we believe that a thorough testing of pipeline performance is necessary to produce reproducible research outcomes, avoiding biased results and allowing for better understanding of human brain function.



NeuroImage 190 (2019) 466–479

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Optimization of preprocessing strategies in Positron Emission Tomography (PET) neuroimaging: A [^{11}C]DASB PET study

Martin Nørgaard[a,b], Melanie Ganz[b,c], Claus Svarer[a], Vibe G. Frokjaer[a], Douglas N. Greve[e], Stephen C. Strother[d], Gitte M. Knudsen[a,b,*]

[a] Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[b] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[c] Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
[d] Rotman Research Institute, Baycrest, Department of Medical Biophysics, University of Toronto, Toronto, Canada
[e] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

Keywords:
Positron emission tomography
Preprocessing
Head motion
Optimization
Partial volume correction
Kinetic modeling
Test-retest
[^{11}C]DASB

ABSTRACT

Positron Emission Tomography (PET) is an important neuroimaging tool to quantify the distribution of specific molecules in the brain. The quantification is based on a series of individually designed data preprocessing steps (pipeline) and an optimal preprocessing strategy is per definition associated with less noise and improved statistical power, potentially allowing for more valid neurobiological interpretations. In spite of this, it is currently unclear how to design the best preprocessing pipeline and to what extent the choice of each preprocessing step in the pipeline minimizes subject-specific errors.

To evaluate the impact of various preprocessing strategies, we systematically examined 384 different pipeline strategies in data from 30 healthy participants scanned twice with the serotonin transporter (5-HTT) radioligand [^{11}C]DASB. Five commonly used preprocessing steps with two to four options were investigated: (1) motion correction (MC) (2) co-registration (3) delineation of volumes of interest (VOIs) (4) partial volume correction (PVC), and (5) kinetic modeling. To quantitatively compare and evaluate the impact of various preprocessing strategies, we used the performance metrics: test-retest bias, within- and between-subject variability, the intraclass-correlation coefficient, and global signal-to-noise ratio. We also performed a power analysis to estimate the required sample size to detect either a 5% or 10% difference in 5-HTT binding as a function of preprocessing pipeline.

The results showed a complex downstream dependency between the various preprocessing steps on the performance metrics. The choice of MC had the most profound effect on 5-HTT binding, prior to the effects caused by PVC and kinetic modeling, and the effects differed across VOIs. Notably, we observed a negative bias in 5-HTT binding across test and retest in 98% of pipelines, ranging from 0 to 6% depending on the pipeline. Optimization of the performance metrics revealed a trade-off in within- and between-subject variability at the group-level with opposite effects (i.e. minimization of within-subject variability increased between-subject variability and vice versa). The sample size required to detect a given effect size was also compromised by the preprocessing strategy, resulting in up to 80% increases in sample size needed to detect a 5% difference in 5-HTT binding.

This is the first study to systematically investigate and demonstrate the effect of choosing different preprocessing strategies on the outcome of dynamic PET studies. We provide a framework to show how optimal and maximally powered neuroimaging results can be obtained by choosing appropriate preprocessing strategies and we provide recommendations depending on the study design.

In addition, the results contribute to a better understanding of methodological uncertainty and variability in preprocessing decisions for future group- and/or longitudinal PET studies.

- PET preprocessing workflow, optimization strategies, variations in preprocessing steps

- Motion Correction
- Co-Registration
- Delineation of Volumes of Interest
- Partial Volume Correction
- Pharmacokinetic Modelling

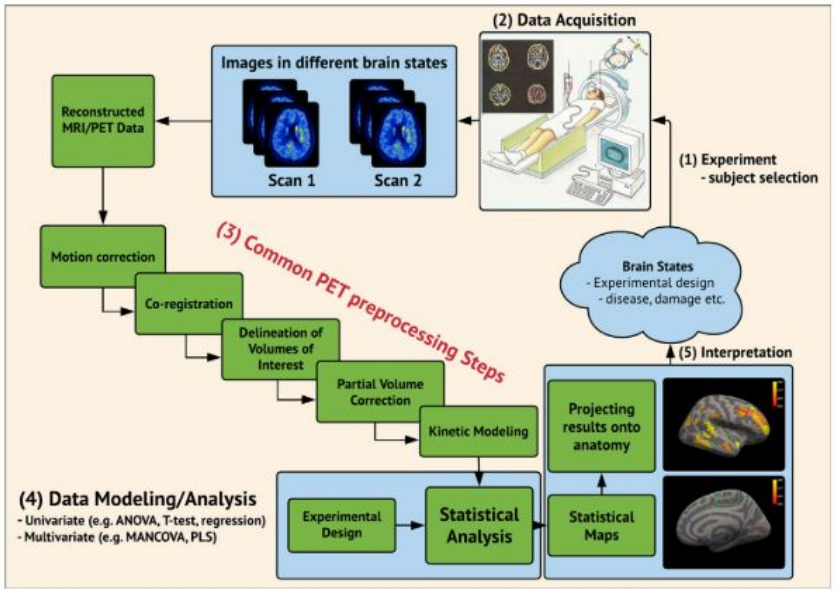


Figure 2: Neuroimaging workflow. Image adapted from [3]

Motion Correction

- Remove motion artefacts due to head movements or respiration.

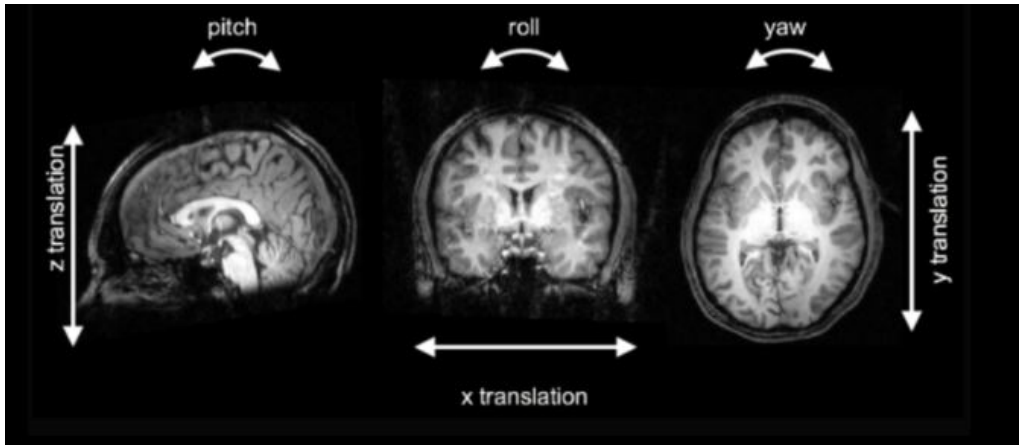


Figure 3: Motion of a rigid body. [a]

[a] http://www.newbi4fmri.com/tutorial-5-motion

Motion Correction

- Remove motion artefacts due to head movements or respiration.
- Mean, sum or median image can be used as a reference image

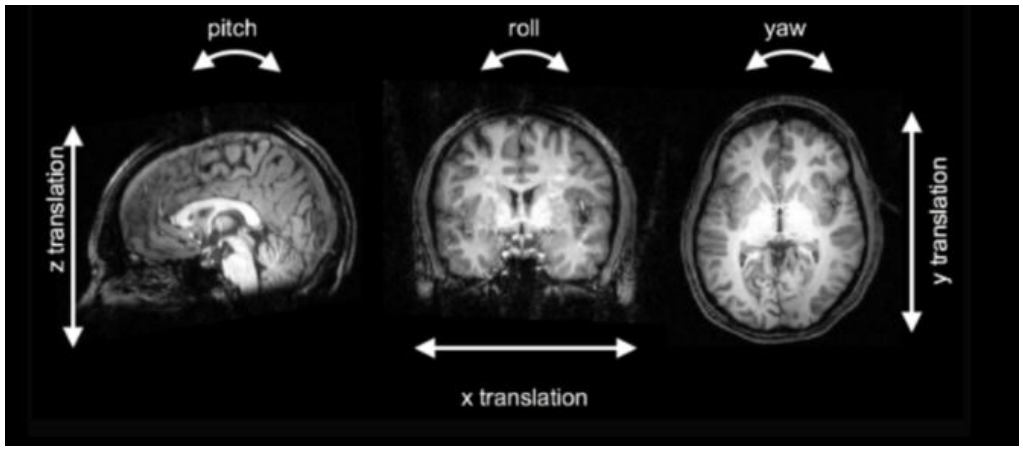


Figure 3: Motion of a rigid body. [a]

[a] http://www.newbi4fmri.com/tutorial-5-motion

Co-Registration

- PET images are studied along with its corresponding anatomical MR image.

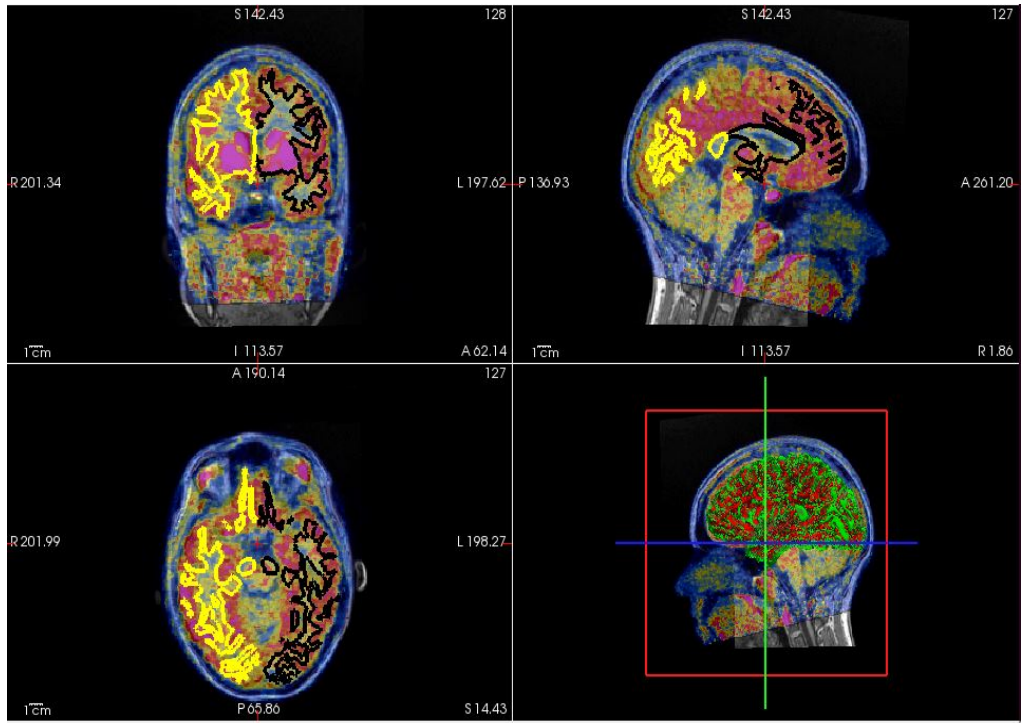


Figure 4: PET-MR coregistration for a subject from the PET-CIMBI dataset

Co-Registration

- PET images are studied along with its corresponding anatomical MR image.
- They need to be co-registered.

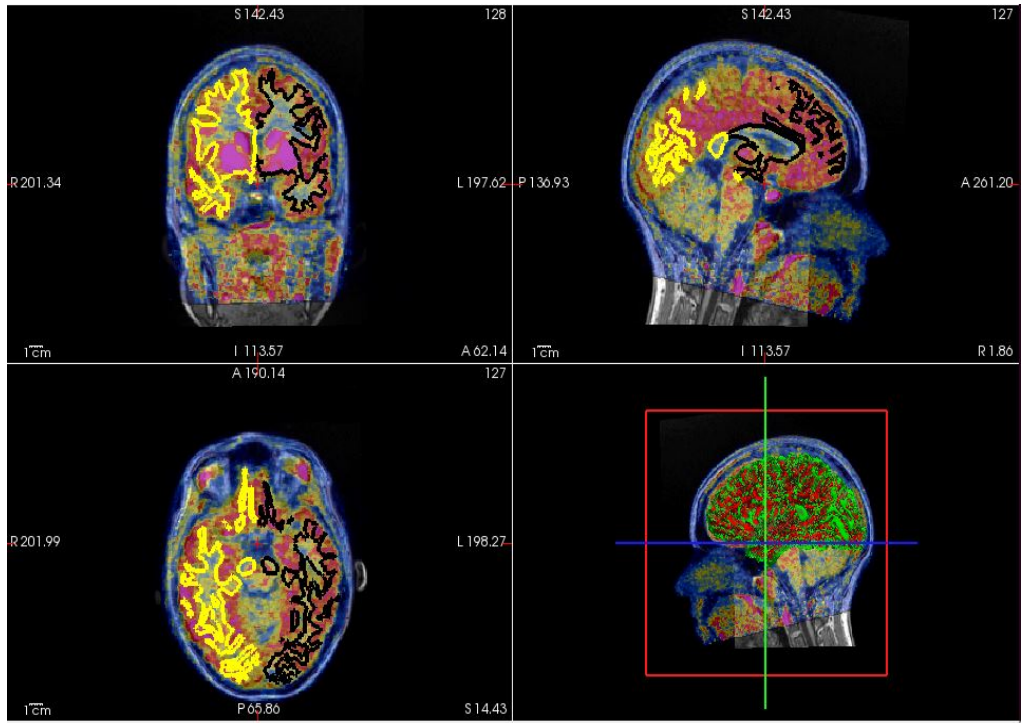


Figure 4: PET-MR coregistration for a subject from the PET-CIMBI dataset

Co-Registration

- PET images are studied along with its corresponding anatomical MR image.
- They need to be co-registered.
- Common techniques include Boundary based registration, normalized mutual information based registration.

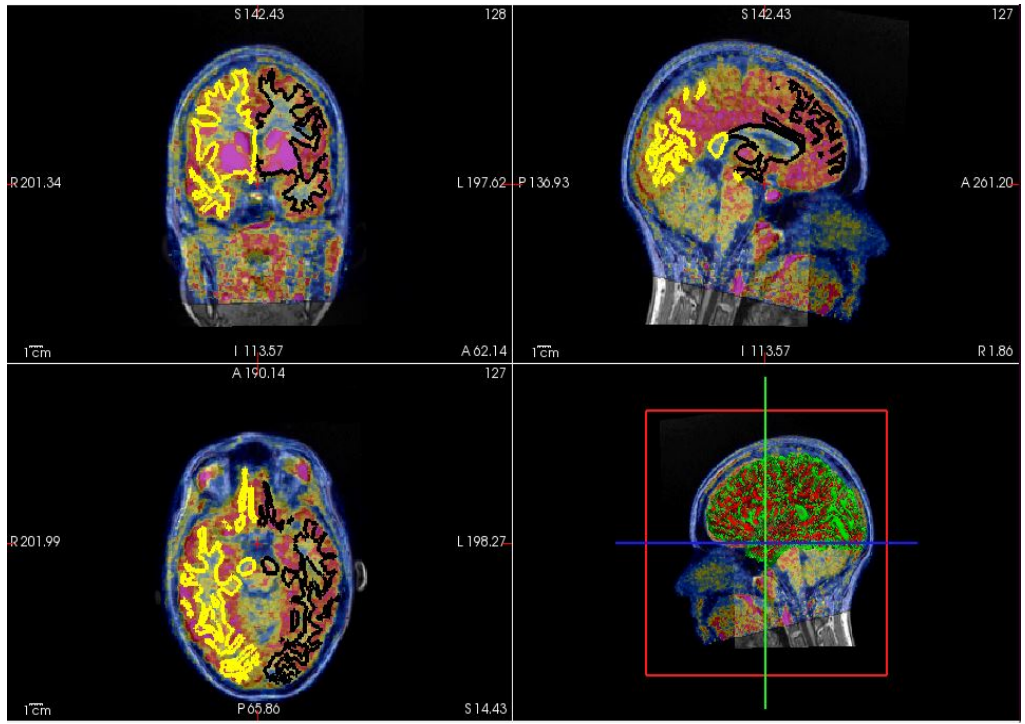


Figure 4: PET-MR coregistration for a subject from the PET-CIMBI dataset

Delineation of Volumes of Interest

- Delineate subset of regions or structures within the brain using predefined brain atlases.

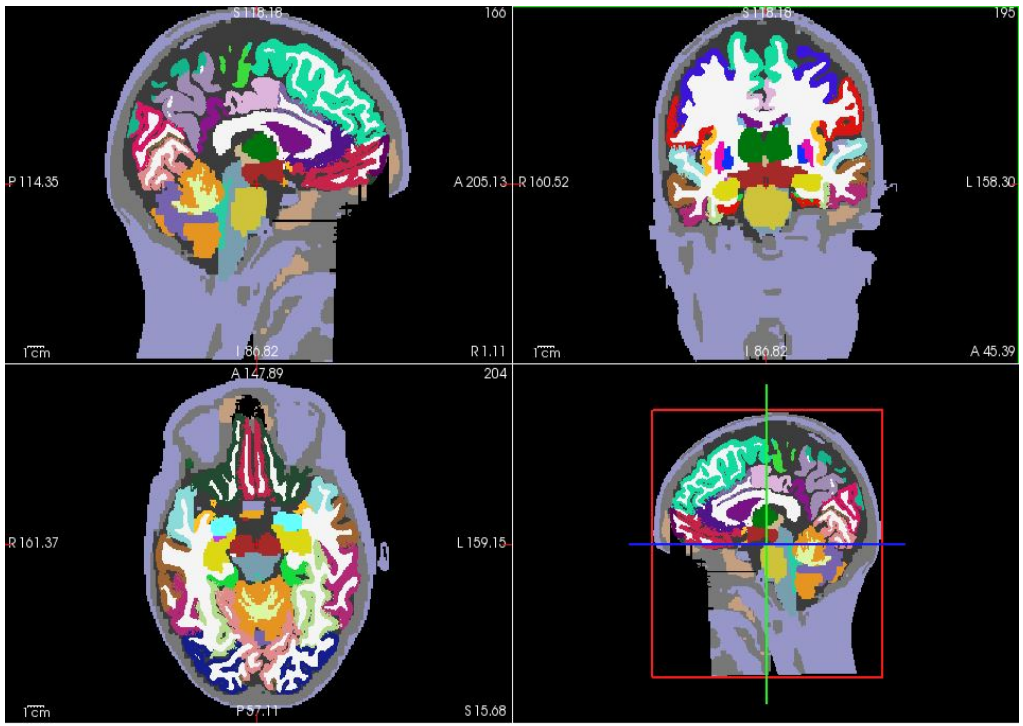


Figure 5: Segmention of brain regions using PETSurfer for an example subject from PET-CIMBI dataset

Delineation of Volumes of Interest

- Delineate subset of regions or structures within the brain using predefined brain atlases.
- PETSurfer provides an implementation for this.

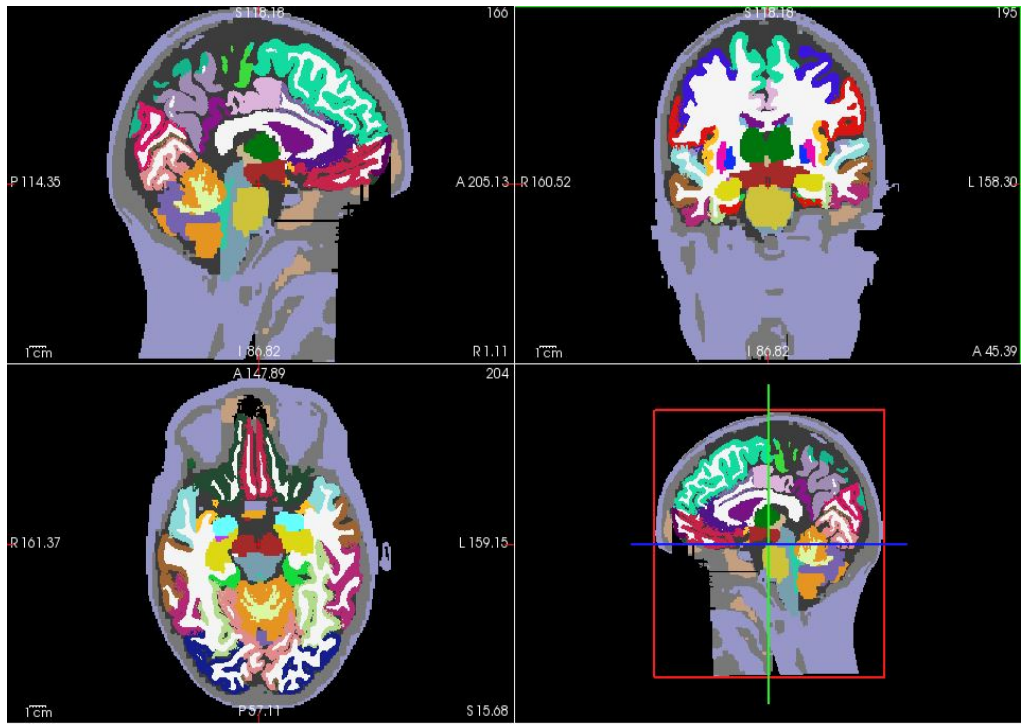
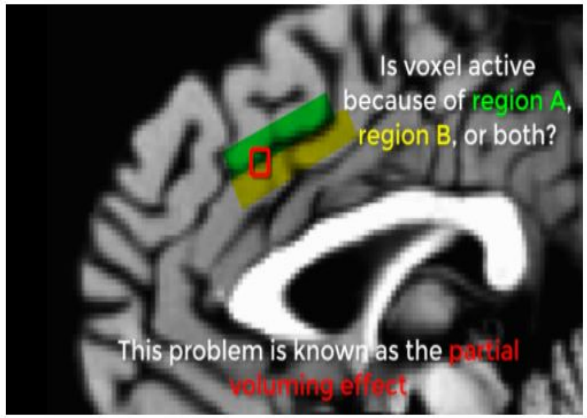


Figure 5: Segmention of brain regions using PETSurfer for an example subject from PET-CIMBI dataset

Partial Volume Correction

- Limited spatial resolution of PET scanners causes Partial Volume Effects.
- These need to be corrected for. Common techniques used for this are Muller Gartner, Geometric transfer Matrix based methods.



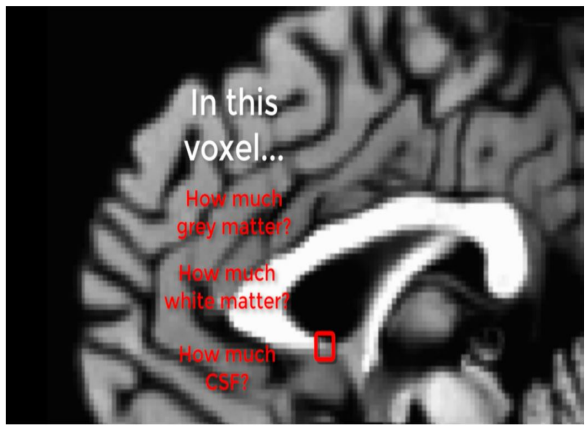


Figure 6: The above figures show partial voluming effects. Images adapted from Andy's Brain Blog

Pharmacokinetic Modelling

- Used to quantify radiotracer binding at the receptor site and output a non-displacable binding potential.

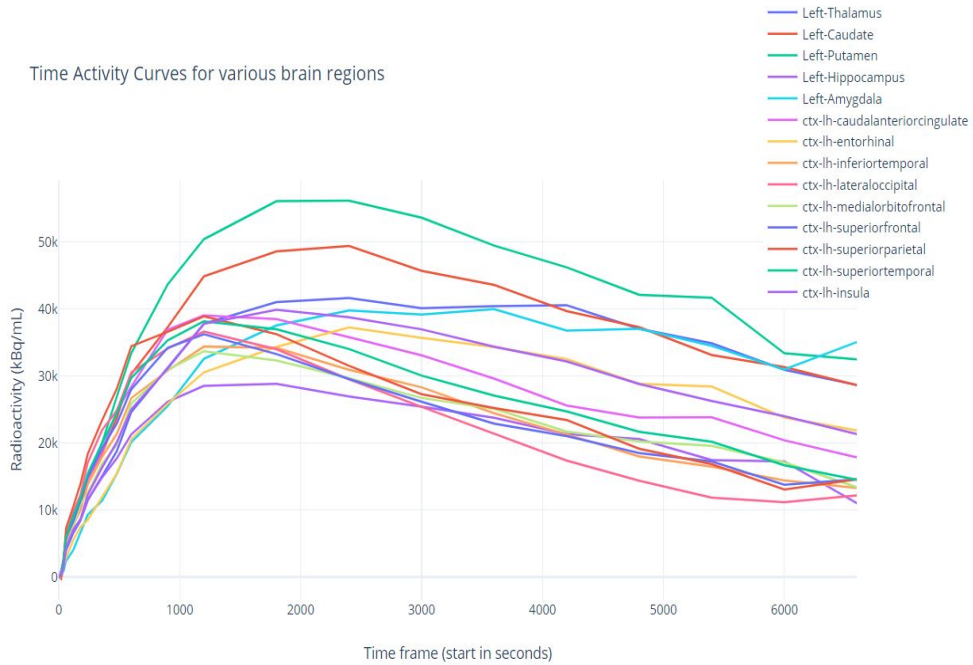


Figure 7: Activity Curves for different brain regions

Pharmacokinetic Modelling

- Used to quantify radiotracer binding at the receptor site and output a non-displacable binding potential.
- Kinetic Models can be fit using Time Activity Curves for different regions to estimate BPnd values

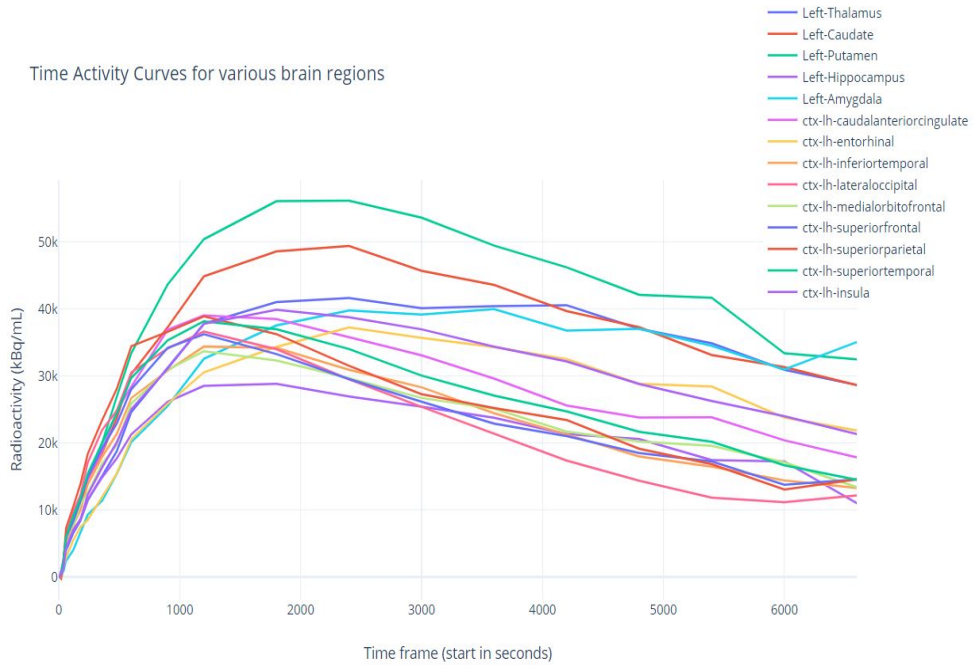


Figure 7: Activity Curves for different brain regions

Pharmacokinetic Modelling

- Used to quantify radiotracer binding at the receptor site and output a non-displacable binding potential.
- Kinetic Models can be fit using Time Activity Curves for different regions to estimate BPnd values
- Some of the kinetic modelling techniques are SRTM, MRTM1, MRTM2. Implementations of these models have been provided in PETSurfer.

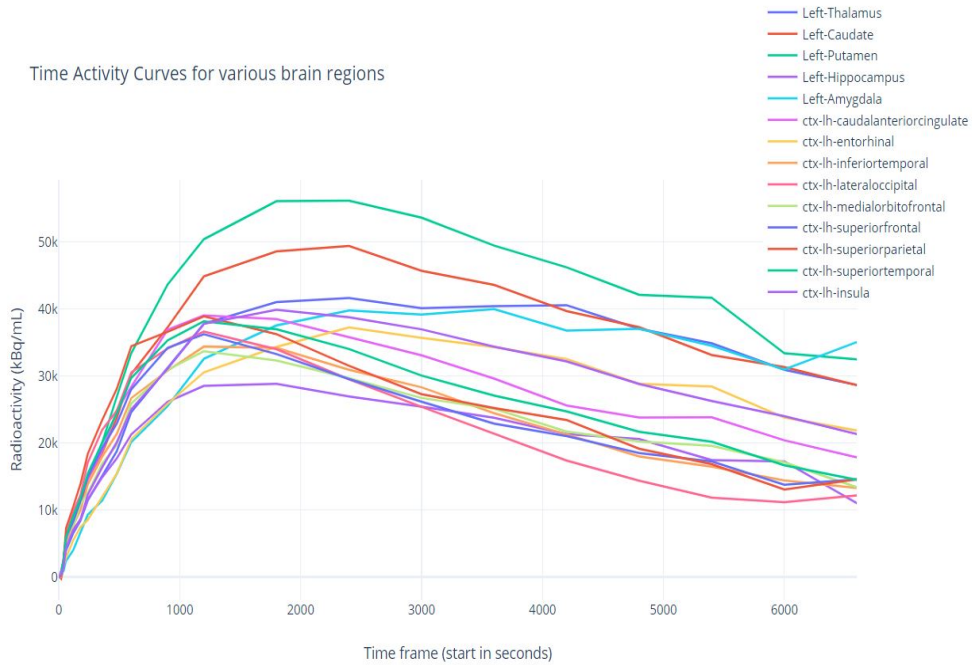


Figure 7: Activity Curves for different brain regions

Reproducibility crisis in the neuroimaging community

Original Article

JCBFM

Journal of Cerebral Blood Flow & Metabolism
2021, Vol. 41(10) 2778–2796

DOI: 10.1177/0271678X211015101
journals.sagepub.com/home/jcbfm
SAGE

Reproducibility of findings in modern PET neuroimaging: insight from the NRM2018 grand challenge

Mattia Veronese[1,*], Gaia Rizzo[2,*], Martin Belzunce[3], Julia Schubert[1], Graham Searle[2], Alex Whittington[2], Ayla Mansur[2,4], Joel Dunn[3,5], Andrew Reader[3] and Roger N Gunn[2,4]; and the Grand Challenge Participants[#]

Abstract

The reproducibility of findings is a compelling methodological problem that the neuroimaging community is facing these days. The lack of standardized pipelines for image processing, quantification and statistics plays a major role in the variability and interpretation of results, even when the same data are analysed. This problem is well-known in MRI studies, where the indisputable value of the method has been complicated by a number of studies that produce discrepant results. However, any research domain with complex data and flexible analytical procedures can experience a similar lack of reproducibility. In this paper we investigate this issue for brain PET imaging. During the 2018 NeuroReceptor Mapping conference, the brain PET community was challenged with a computational contest involving a simulated neurotransmitter release experiment. Fourteen international teams analysed the same imaging dataset, for which the ground-truth was known. Despite a plurality of methods, the solutions were consistent across participants, although not identical. These results should create awareness that the increased sharing of PET data alone will only be one component of enhancing confidence in neuroimaging results and that it will be important to complement this with full details of the analysis pipelines and procedures that have been used to quantify data.

Keywords

PET, data analysis, data sharing, reproducibility crisis, "NRM2018 PET Grand Challenge"

Received 23 December 2020; Revised 10 March 2021; Accepted 3 April 2021

Review › Nat Rev Neurosci. 2017 Feb;18(2):115-126. doi: 10.1038/nrn.2016.167. Epub 2017 Jan 5.

Scanning the horizon: towards transparent and reproducible neuroimaging research

Russell A Poldrack [1], Chris I Baker [2], Joke Durnez [1 3], Krzysztof J Gorgolewski [1], Paul M Matthews [4], Marcus R Munafò [5 6], Thomas E Nichols [7], Jean-Baptiste Poline [8], Edward Vul [9], Tal Yarkoni [10]

Affiliations + expand

PMID: 28053326 PMCID: PMC6910649 DOI: 10.1038/nrn.2016.167

Free PMC article

Abstract

Functional neuroimaging techniques have transformed our ability to probe the neurobiological basis of behaviour and are increasingly being applied by the wider neuroscience community. However, concerns have recently been raised that the conclusions that are drawn from some human neuroimaging studies are either spurious or not generalizable. Problems such as low statistical power, flexibility in data analysis, software errors and a lack of direct replication apply to many fields, but perhaps particularly to functional MRI. Here, we discuss these problems, outline current and suggested best practices, and describe how we think the field should evolve to produce the most meaningful and reliable answers to neuroscientific questions.

- Concerns have been raised over the results published by neuroimaging studies not being reproducible and generalizable.

- Concerns have been raised over the results published by neuroimaging studies not being reproducible and generalizable.
- This may be attributed to lack of standardized pipelines for processing and analysis of neuroimaging data.

- Concerns have been raised over the results published by neuroimaging studies not being reproducible and generalizable.
- This may be attributed to lack of standardized pipelines for processing and analysis of neuroimaging data.
- In order to enhance trustworthiness in the results published by various neuroimaging studies, it is important to promote data and code sharing.

- Concerns have been raised over the results published by neuroimaging studies not being reproducible and generalizable.
- This may be attributed to lack of standardized pipelines for processing and analysis of neuroimaging data.
- In order to enhance trustworthiness in the results published by various neuroimaging studies, it is important to promote data and code sharing.
- To promote transparency in methodology it is also essential to provide full details of the pipelines used for analysis.

How do we define reproducibility and robustness?

- **Computational Reproducibility**: When detailed information is provided about *code*, *software*, *hardware* and *implementation details*, the results from the preprocessing pipeline should be consistent across different *computational environments*.

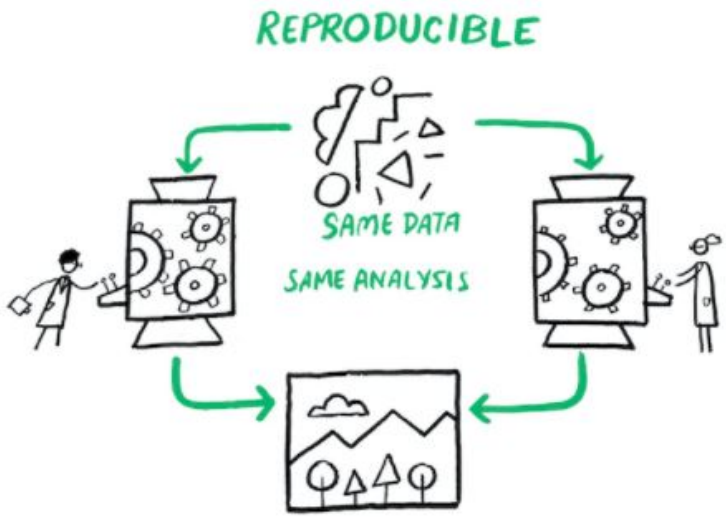


Figure 8: Image adapted from The Turing Way handbook. DOI: 10.5281/zenodo.3332807

How do we define reproducibility and robustness?

- **Computational Reproducibility**: When detailed information is provided about *code*, *software*, *hardware* and *implementation details*, the results from the preprocessing pipeline should be consistent across different *computational environments*.
- **Robustness**: The preprocessing pipeline should be *robust to errors* and should be able to run successfully on data coming from different different radiotracers and PET scanners.

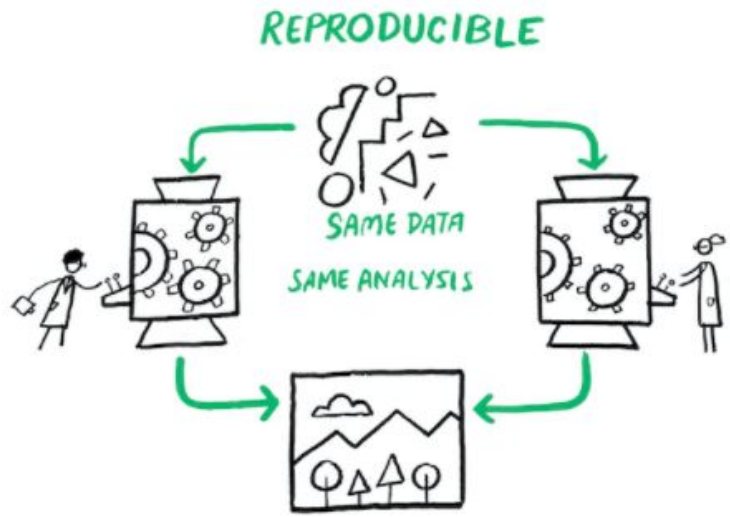


Figure 8: Image adapted from The Turing Way handbook. DOI: 10.5281/zenodo.3332807

Figure 9: Reproducible Research. Image adapted from *The Turing Way handbook* DOI: 10.5281/zenodo.3332807
https://the-turing-way.netlify.app/welcome.html

Objectives

- Development of a preprocessing pipeline involving *motion correction*, *co-registration*, *delineation of volumes of interest* and *partial volume correction* and *pharmacokinetic modelling*.

Objectives

- Development of a preprocessing pipeline involving *motion correction*, *co-registration*, *delineation of volumes of interest and partial volume correction* and *pharmacokinetic modelling*.
- Integration of all these into a single workflow implemented in Python.

Objectives

- Development of a preprocessing pipeline involving *motion correction*, *co-registration*, *delineation of volumes of interest and partial volume correction* and *pharmacokinetic modelling*.
- Integration of all these into a single workflow implemented in Python.
- Testing and evaluation of the pipeline against several datasets coming from different radiotracers and PET scanners.

Objectives

- Development of a preprocessing pipeline involving *motion correction*, *co-registration*, *delineation of volumes of interest and partial volume correction* and *pharmacokinetic modelling*.
- Integration of all these into a single workflow implemented in Python.
- Testing and evaluation of the pipeline against several datasets coming from different radiotracers and PET scanners.
- Releasing the pipeline to the community as open source code and fully executable in a docker container.

Dataset

- **[11C]SB207145 PET CIMBI [1]** (1 subject, 1 baseline & 1 rescan session for each subject)
- **[11C]DASB PET CIMBI [2]** (2 subjects, 1 baseline & 1 rescan session for each subject)
- **[11C]DASB 5-HTT [4]** (16 subjects, 1 baseline & 1 rescan session for each subject)

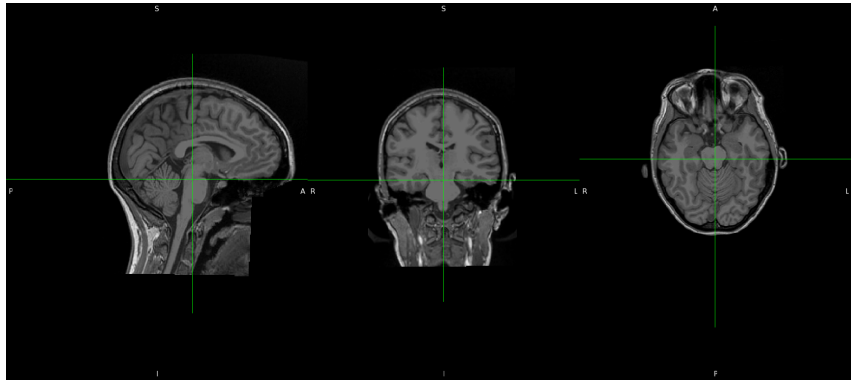


Figure 10: T1 weighted MR image for the baseline session for a subject from the PET CIMBI dataset.

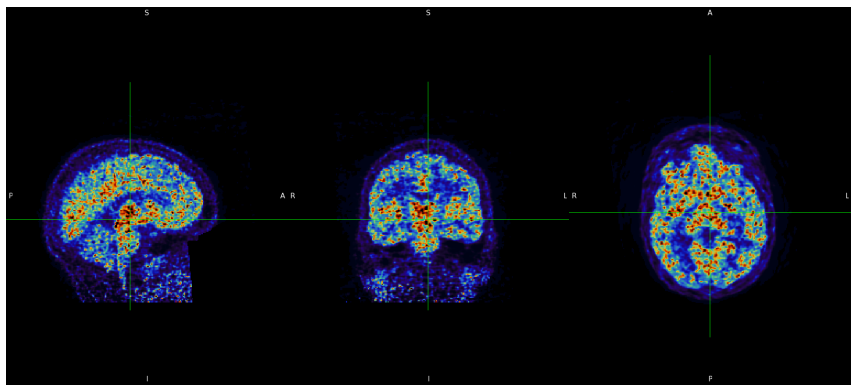


Figure 11: PET image for the baseline session for a subject from the PET CIMBI dataset.

Input Structure: BIDS[1]

```
PET_CIMBI
├─ dataset_description.json
├─ participants.json
├─ participants.tsv
├─ README
└─ sub-01
   ├─ ses-baseline
   │  ├─ anat
   │  │  ├─ sub-01_ses-baseline_T1w.json
   │  │  └─ sub-01_ses-baseline_T1w.nii
   │  └─ pet
   │     ├─ sub-01_ses-baseline_pet.json
   │     └─ sub-01_ses-baseline_pet.nii.gz
   └─ ses-rescan
      ├─ anat
      │  ├─ sub-01_ses-rescan_T1w.json
      │  └─ sub-01_ses-rescan_T1w.nii
      └─ pet
         ├─ sub-01_ses-rescan_pet.json
         └─ sub-01_ses-rescan_pet.nii.gz
```

Open Access | Published: 21 June 2016

The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments

Krzysztof J. Gorgolewski, Tibor Auer, ... Russell A. Poldrack + Show authors

Scientific Data **3**, Article number: 160044 (2016) | Cite this article
44k Accesses | **336** Citations | **107** Altmetric | Metrics

Abstract

The development of magnetic resonance imaging (MRI) techniques has defined modern neuroimaging. Since its inception, tens of thousands of studies using techniques such as functional MRI and diffusion weighted imaging have allowed for the non-invasive study of the brain. Despite the fact that MRI is routinely used to obtain data for neuroscience research, there has been no widely adopted standard for organizing and describing the data collected in an imaging experiment. This renders sharing and reusing data (within or between labs) difficult if not impossible and unnecessarily complicates the application of automatic pipelines and quality assurance protocols. To solve this problem, we have developed the Brain Imaging Data Structure (BIDS), a standard for organizing and describing MRI datasets. The BIDS standard uses file formats compatible with existing software, unifies the majority of practices already common in the field, and captures the metadata necessary for most common data processing operations.

PET-BIDS, an extension to the brain imaging data structure for positron emission tomography

Martin Norgaard[1,2], Granville J. Matheson[3,4], Hanne D. Hansen[1,5], Adam Thomas[6], Graham Searle[7], Gaia Rizzo[7], Mattia Veronese[8,9], Alessio Giacomel[8], Maqsood Yaqub[10], Matteo Tonietto[11], Thomas Funck[12], Ashley Gillman[13], Hugo Boniface[14], Alexandre Routier[15], Jelle R. Dalenberg[16], Tobey Betthauser[17], Franklin Feingold[2], Christopher J. Markiewicz[2], Krzysztof J. Gorgolewski[2], Ross W. Blair[2], Stefan Appelhoff[18], Remi Gau[19], Taylor Salo[20], Guiomar Niso[21], Cyril Pernet[1], Christophe Phillips[22], Robert Oostenveld[23,24], Jean-Dominique Gallezot[25], Richard E. Carson[25], Gitte M. Knudsen[1], Robert B. Innis[26], and Melanie Ganz[1,27,*]

[1] https://bids-specification.readthedocs.io/en/stable/

Building the pipeline: Workflow Design and Architecture

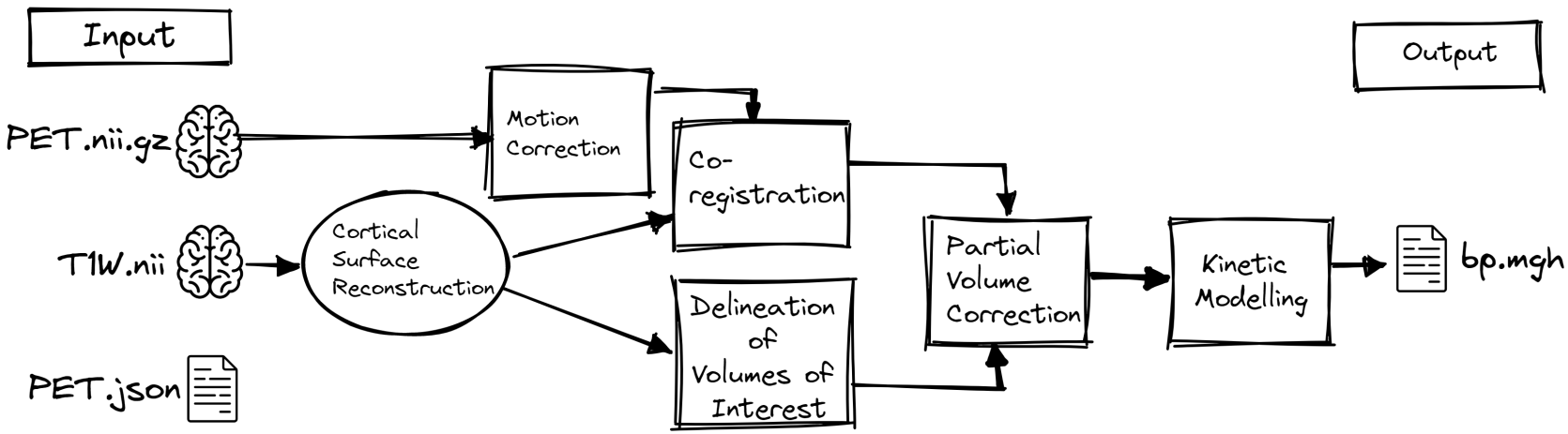


Figure 12: Rough Sketch of the Pipeline

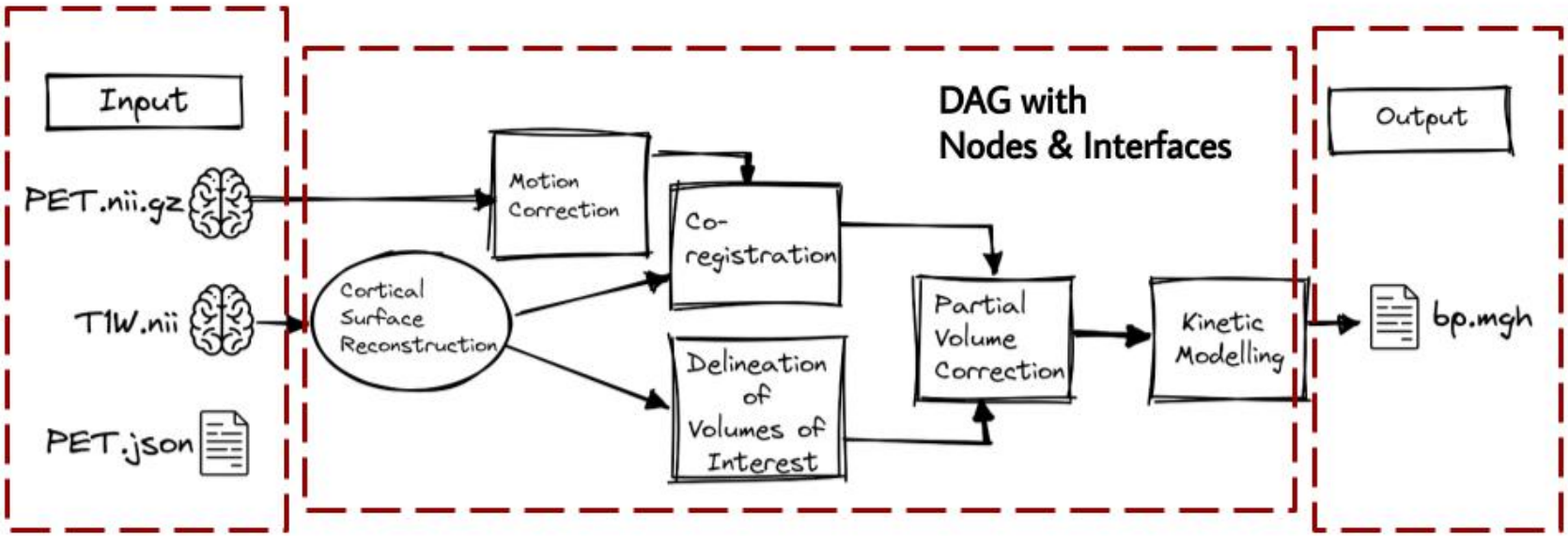
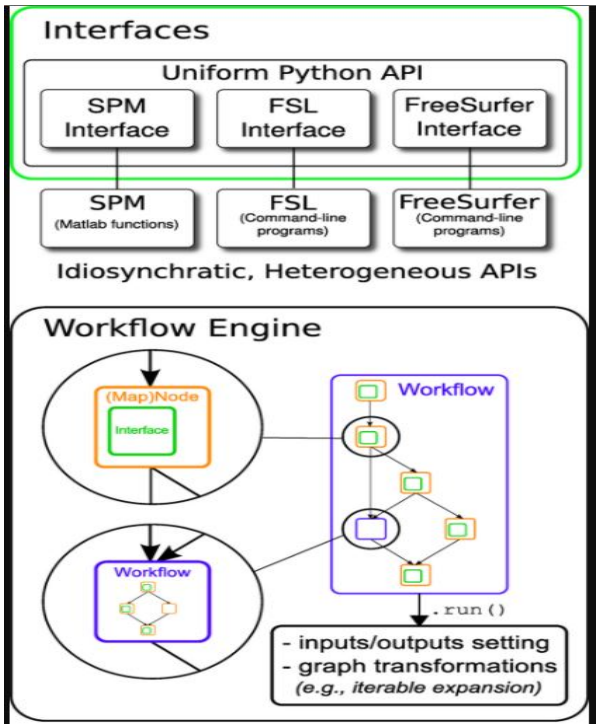


Figure 13: Visualizing the pipeline as a Directed Acyclic Graph

Building the preprocessing pipeline using Nipype

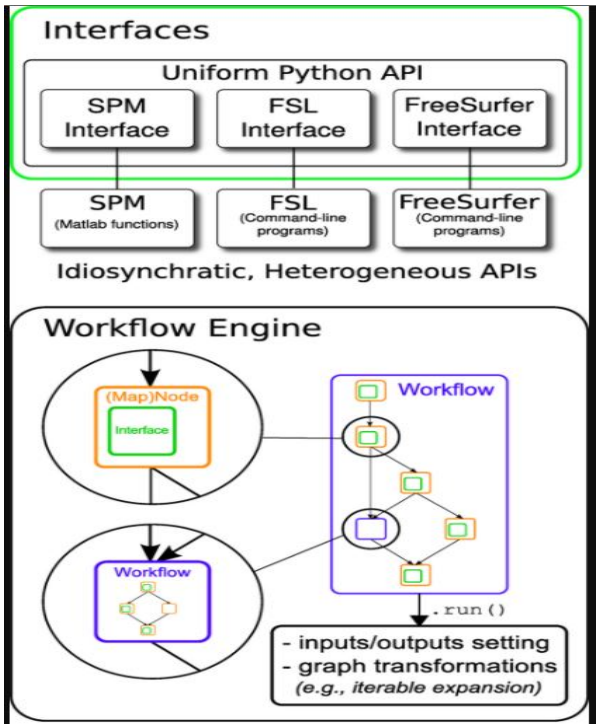


- Nodes

Node	Functional Interface
Cortical Reconstruction	FreeSurfer Recon-All
Motion Correction	FSL MCFlirt
Co-registration	Freesurfer MRI Coreg
Delineation of Volumes of Interest	PETSurfer GTMSeg
Partial Volume Correction	PETSurfer GTMPVC
Kinetic Modelling (MRTM1)	PETSurfer MRTM1
Kinetic Modelling (MRTM2)	PETSurfer MRTM2

Table 1: Summary of the interfaces & inputs for the PETPipeline

Building the preprocessing pipeline using Nipype

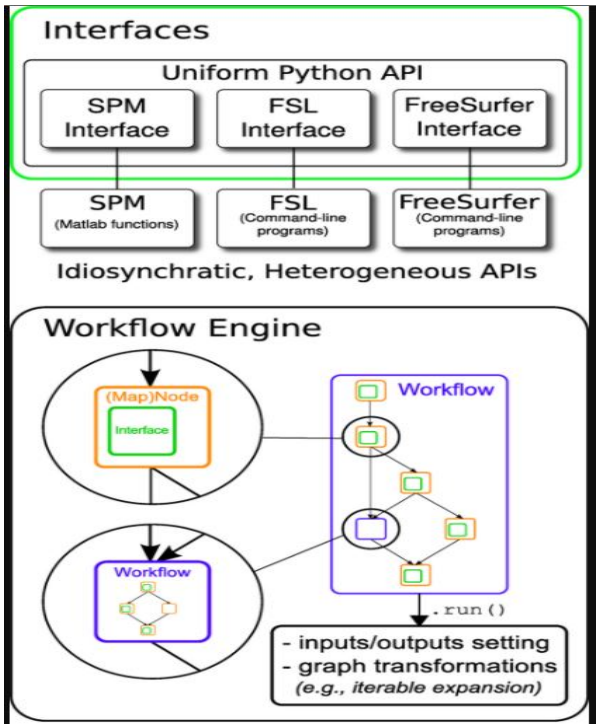


- Nodes
- Interfaces

Node	Functional Interface
Cortical Reconstruction	FreeSurfer Recon-All
Motion Correction	FSL MCFlirt
Co-registration	Freesurfer MRI Coreg
Delineation of Volumes of Interest	PETSurfer GTMSeg
Partial Volume Correction	PETSurfer GTMPVC
Kinetic Modelling (MRTM1)	PETSurfer MRTM1
Kinetic Modelling (MRTM2)	PETSurfer MRTM2

Table 1: Summary of the interfaces & inputs for the PETPipeline

Building the preprocessing pipeline using Nipype

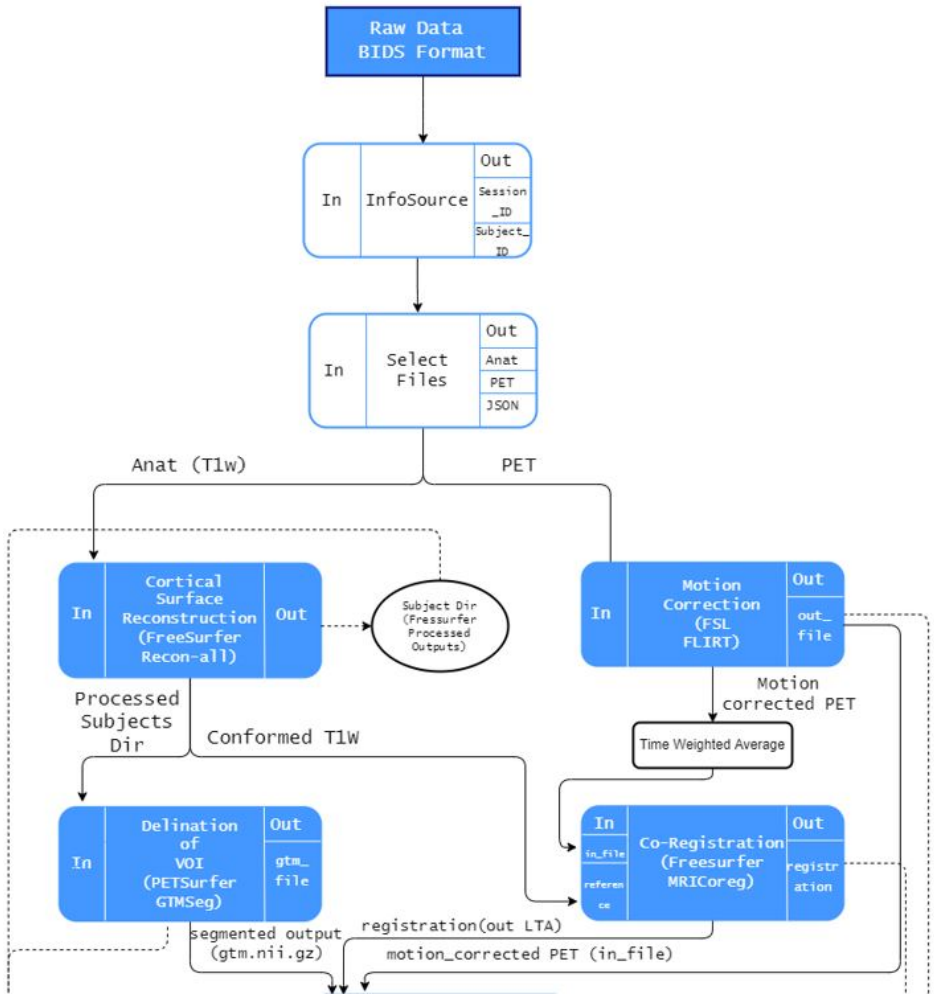


- Nodes
- Interfaces
- Workflows

Node	Functional Interface
Cortical Reconstruction	FreeSurfer Recon-All
Motion Correction	FSL MCFlirt
Co-registration	Freesurfer MRI Coreg
Delineation of Volumes of Interest	PETSurfer GTMSeg
Partial Volume Correction	PETSurfer GTMPVC
Kinetic Modelling (MRTM1)	PETSurfer MRTM1
Kinetic Modelling (MRTM2)	PETSurfer MRTM2

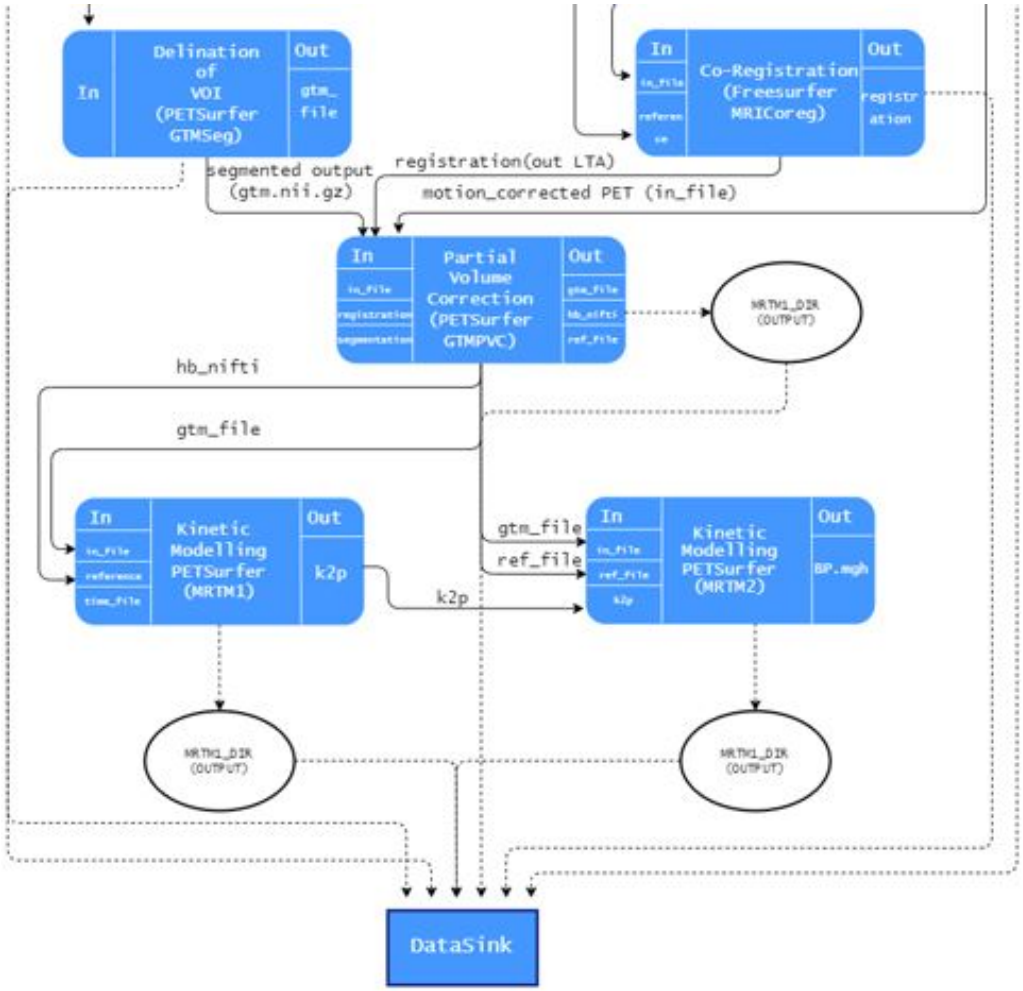
Table 1: Summary of the interfaces & inputs for the PETPipeline

Interaction of nodes & interfaces in PETPipeline I



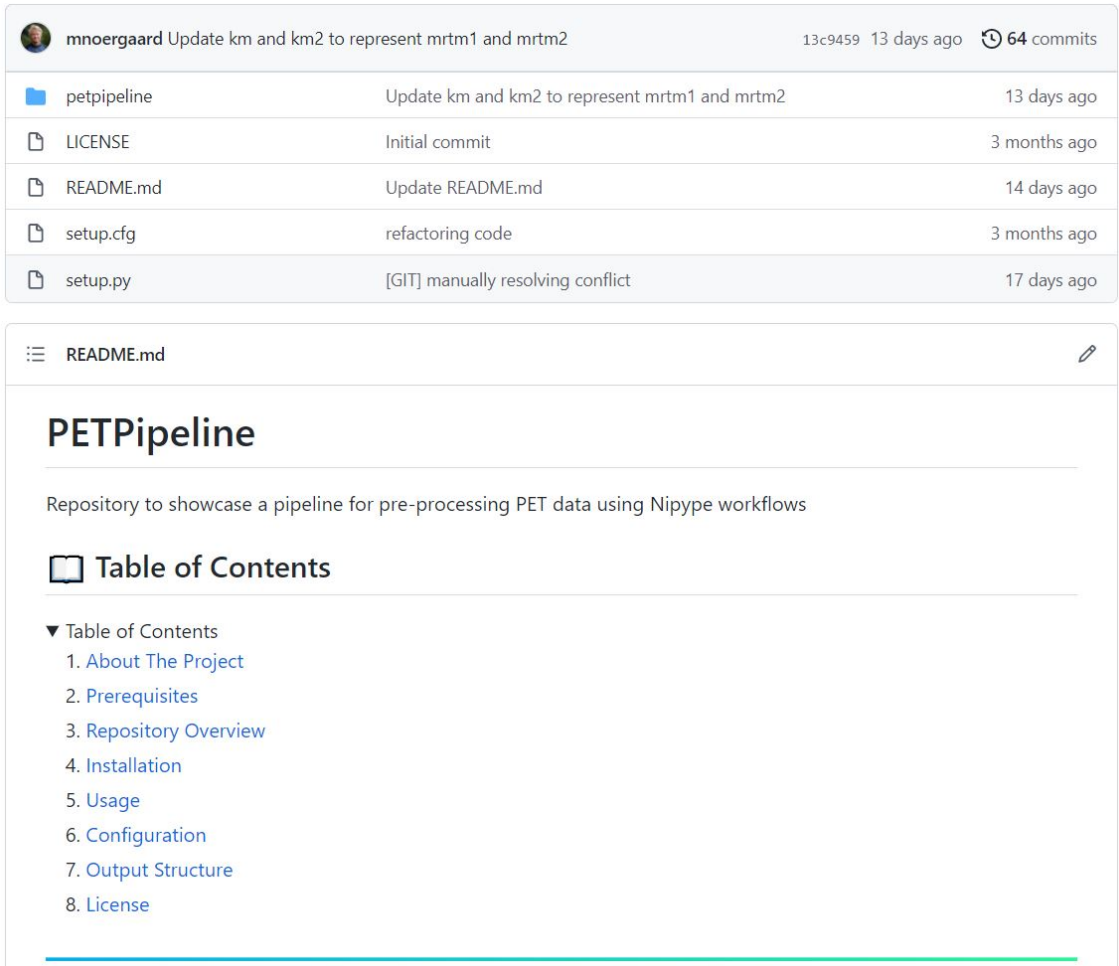
https://drive.google.com/file/d/1pUOebAfr9Y2WmsP0LOwO-XVkduORX-IR/view?usp=sharing

Interaction of nodes & interfaces in PETPipeline II



https://drive.google.com/file/d/1pUOebAfr9Y2WmsP0LOwO-XVkduORX-IR/view?usp=sharing

Final Outcome: PETPipeline



mnoergaard Update km and km2 to represent mrtm1 and mrtm2 — 13c9459 — 13 days ago — **64** commits

petpipeline	Update km and km2 to represent mrtm1 and mrtm2	13 days ago
LICENSE	Initial commit	3 months ago
README.md	Update README.md	14 days ago
setup.cfg	refactoring code	3 months ago
setup.py	[GIT] manually resolving conflict	17 days ago

README.md

PETPipeline

Repository to showcase a pipeline for pre-processing PET data using Nipype workflows

Table of Contents

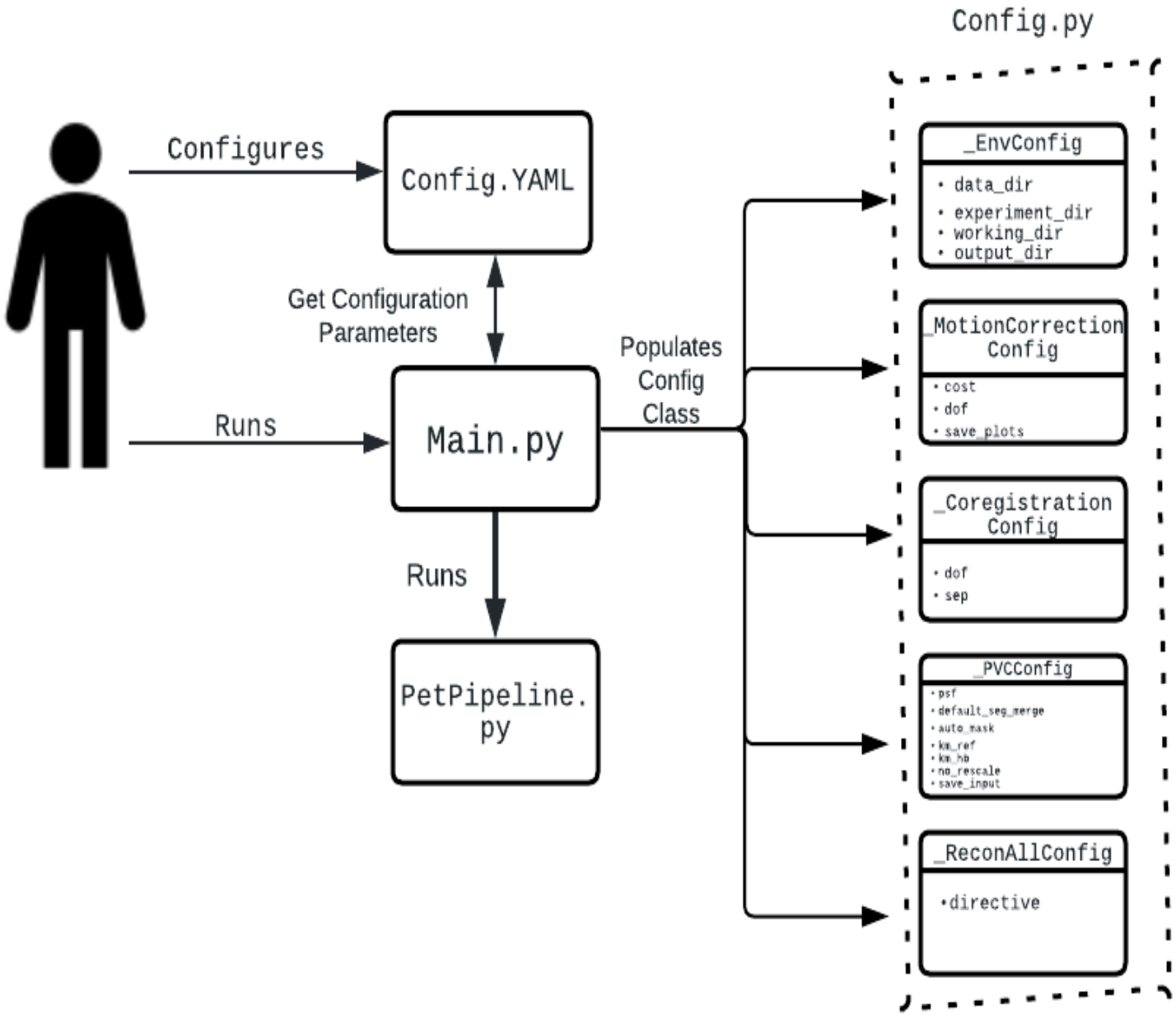
▼ Table of Contents

1. About The Project
2. Prerequisites
3. Repository Overview
4. Installation
5. Usage
6. Configuration
7. Output Structure
8. License

Figure 15: Overview of the PETPipeline repository on Github

https://github.com/openneuropet/petpipeline

User Interaction with PETPipeline



Computational Environments for Testing

A computational environment here broadly refers to the system where a particular piece of code or an application is run. It consists of *hardware features (CPUs, cores etc.)* and *software features such as OS, software installed (including their version and configuration), programming language* etc.

Name	OS	CPU	N_Processors	RAM
VM (Virtual Box)	Ubuntu 18.04	Intel i7-7700 2.8 Ghz	2 (1)	10 GB
NRU Server	openSuse Leap 15.3	Intel(R) Xeon(R) 3.00 Ghz	72 (used 30)	750 GB
Docker 1	Ubuntu 20.04	Intel i7-7700 2.8 GHz	2 (1)	10 GB
Docker 2	Ubuntu 20.04	(QC Intel i7 2.2 GHz)	(1)	12 GB

Table 2: Summary of computational environments for processing of the data using PETPipeline

Evaluation

- 14 brain regions were chosen for analysis namely amygdala, thalamus, putamen, caudate, anterior cingulate cortex, hippocampus, frontal cortex, occipital cortex, temporal cortex, parietal cortex, entorhinal cortex based on previous studies by Nørgaard et al. [3]

Evaluation

- 14 brain regions were chosen for analysis namely amygdala, thalamus, putamen, caudate, anterior cingulate cortex, hippocampus, frontal cortex, occipital cortex, temporal cortex, parietal cortex, entorhinal cortex based on previous studies by Nørgaard et al. [3]
- To test for differences in binding potential (BP) across brain regions and between different computational environments, we did a one way ANOVA, and also compared the regression slopes for statistical differences.

Evaluation

- 14 brain regions were chosen for analysis namely amygdala, thalamus, putamen, caudate, anterior cingulate cortex, hippocampus, frontal cortex, occipital cortex, temporal cortex, parietal cortex, entorhinal cortex based on previous studies by Nørgaard et al. [3]
- To test for differences in binding potential (BP) across brain regions and between different computational environments, we did a one way ANOVA, and also compared the regression slopes for statistical differences.
- The test-retest bias metric [3] was computed between the resulting BP estimates from kinetic modelling obtained as a result of running the pipeline across various computational environments to identify any differences.

$$Bias_{i,j} = \frac{retest_{i,j} - test_{i,j}}{test_{i,j}} * 100$$

Do we see any differences?

Regions	VM	Docker1	Server	Docker2	Mean	Variance
Amygdala	1.66547	1.663995	1.634389	1.663995	1.656962	0.000227
Caudal Anterior Cingulate Cortex	0.707916	0.708198	0.714869	0.708198	0.709795	1.1e-05
Caudate	1.505885	1.507484	1.472801	1.507484	1.498413	0.000292
Entorhinal Cortex	1.008542	1.008232	0.946286	1.008232	0.992823	0.000963
Hippocampus	0.326675	0.324846	0.343451	0.324846	0.329954	8.2e-05
Inferior Temporal Cortex	0.395526	0.39625	0.371875	0.39625	0.389975	0.000146
Insula	0.966093	0.967939	0.928386	0.967939	0.957589	0.00038
Medial Orbitofrontal Cortex	0.421494	0.423251	0.501818	0.423251	0.442454	0.001567
Occipital Cortex	0.284816	0.282593	0.28846	0.282593	0.284616	8e-06
Putamen	1.896689	1.896622	1.880213	1.896622	1.892536	6.7e-05
Superior Frontal Cortex	0.377558	0.376259	0.374473	0.376259	0.376137	2e-06
Superior Parietal Cortex	0.565708	0.566033	0.563758	0.566033	0.565383	1e-06
Superior Temporal Cortex	0.465539	0.461582	0.470955	0.461582	0.464914	2e-05
Thalamus	1.48164	1.483333	1.451222	1.483333	1.474882	0.000249

Table 3: Mean BP values for 14 different brain regions and across 4 different computational environments for a single subject from the [11C]DASB 5-HTT dataset

Do we see any differences?

Regions	VM	Docker1	Server	Docker2	Mean	Variance
Amygdala	1.66547	1.663995	1.634389	1.663995	1.656962	0.000227
Caudal Anterior Cingulate Cortex	0.707916	0.708198	0.714869	0.708198	0.709795	1.1e-05
Caudate	1.505885	1.507484	1.472801	1.507484	1.498413	0.000292
Entorhinal Cortex	1.008542	1.008232	0.946286	1.008232	0.992823	0.000963
Hippocampus	0.326675	0.324846	0.343451	0.324846	0.329954	8.2e-05
Inferior Temporal Cortex	0.395526	0.39625	0.371875	0.39625	0.389975	0.000146
Insula	0.966093	0.967939	0.928386	0.967939	0.957589	0.00038
Medial Orbitofrontal Cortex	0.421494	0.423251	0.501818	0.423251	0.442454	0.001567
Occipital Cortex	0.284816	0.282593	0.28846	0.282593	0.284616	8e-06
Putamen	1.896689	1.896622	1.880213	1.896622	1.892536	6.7e-05
Superior Frontal Cortex	0.377558	0.376259	0.374473	0.376259	0.376137	2e-06
Superior Parietal Cortex	0.565708	0.566033	0.563758	0.566033	0.565383	1e-06
Superior Temporal Cortex	0.465539	0.461582	0.470955	0.461582	0.464914	2e-05
Thalamus	1.48164	1.483333	1.451222	1.483333	1.474882	0.000249

Table 4: Mean BP values for 14 different brain regions and across 4 different computational environments for a single subject from the [11C]DASB 5-HTT dataset

Differences across computational environments

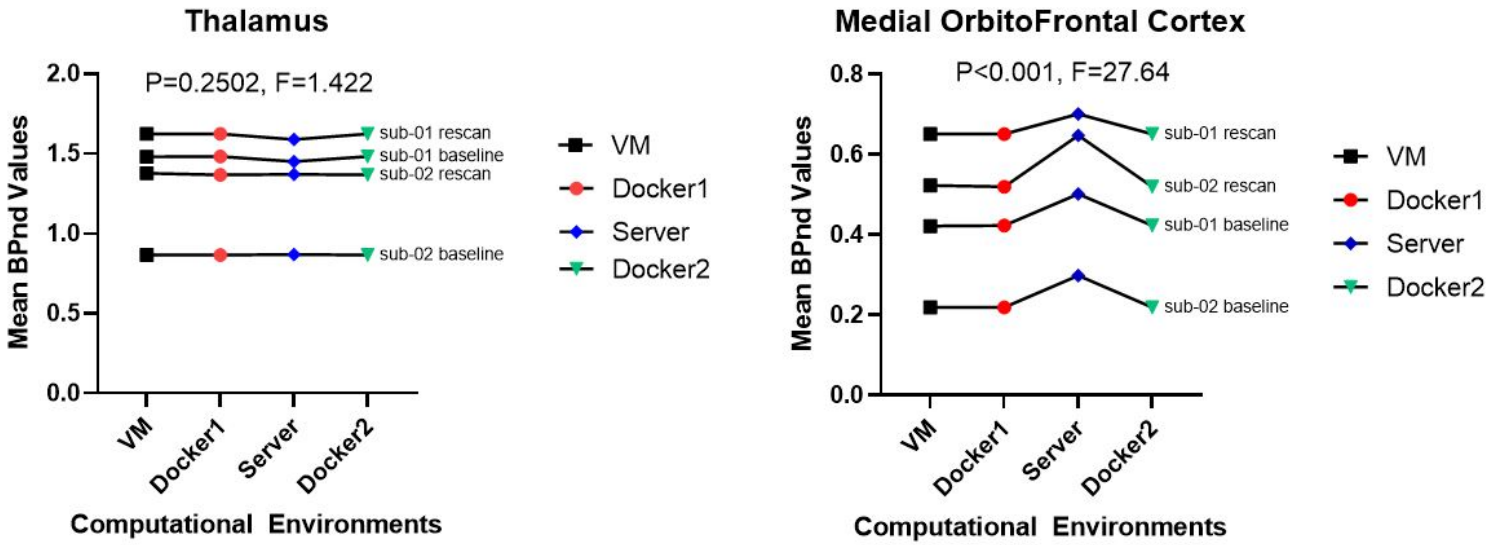


Figure 16: Mean BP estimates across 2 subjects and 2 sessions for different environments for two brain regions from the [11C]DASB 5-HTT dataset.

Differences across subjects

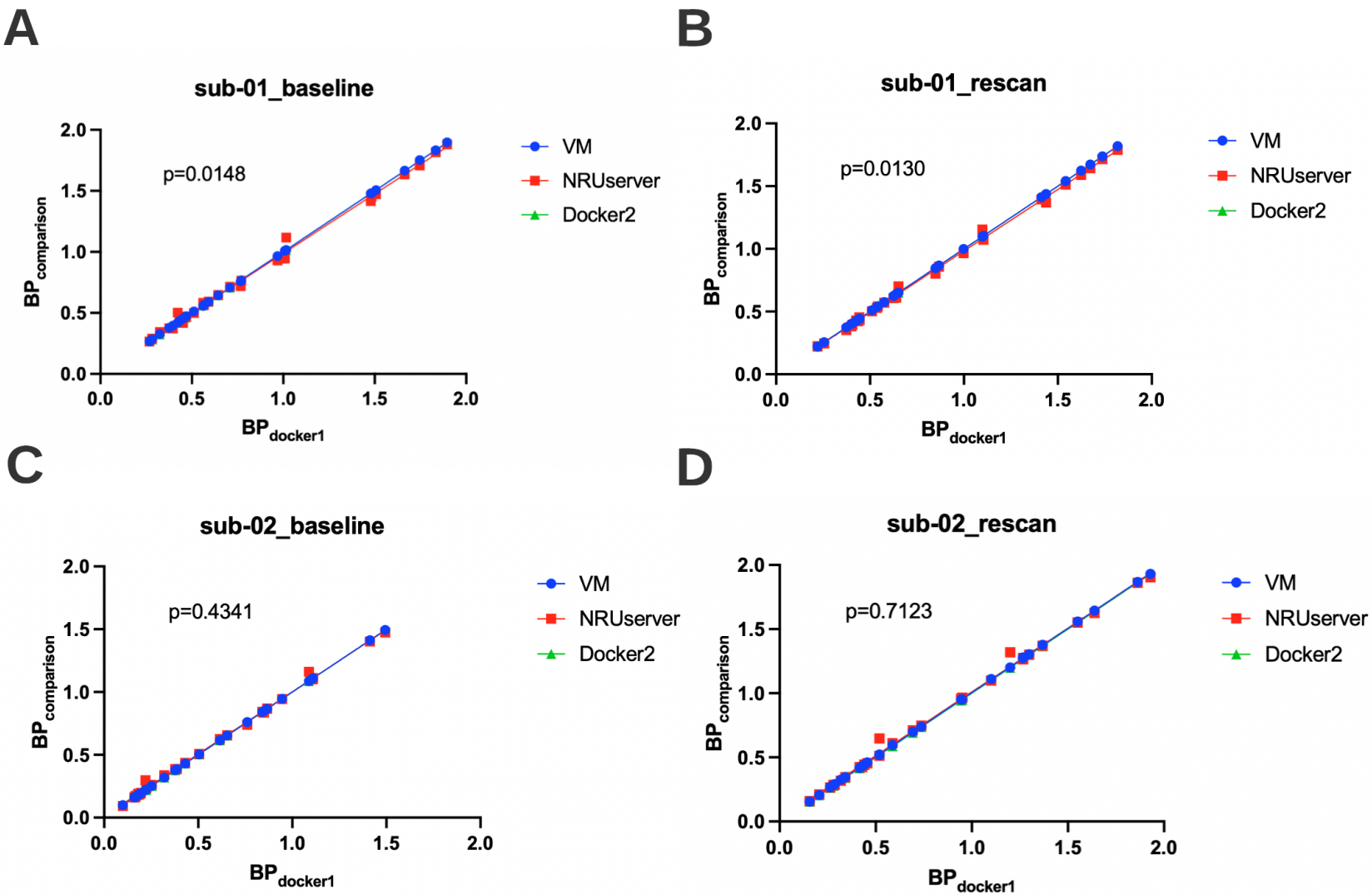


Figure 17: Slope estimates of BP across different brain regions for different subjects and sessions across different computational environments.

Why do these differences arise?

- *significant bias* across *brain regions* and variance in the bias across *subjects*

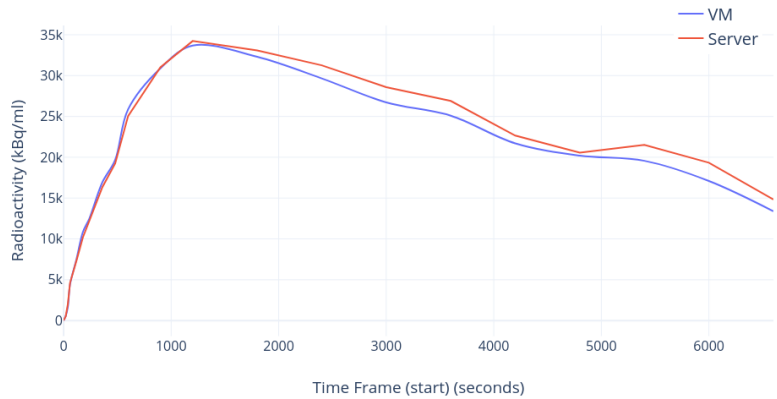


Figure 18: Time activity curve for the Medial Orbitofrontal Cortex across two different computational environments (VM and server), highlighting the uptake of radioactivity within this region as a function of time

Why do these differences arise?

- *significant bias* across *brain regions* and variance in the bias across *subjects*
- Variations due to difference in the cortical and sub-cortical segmentations obtained.

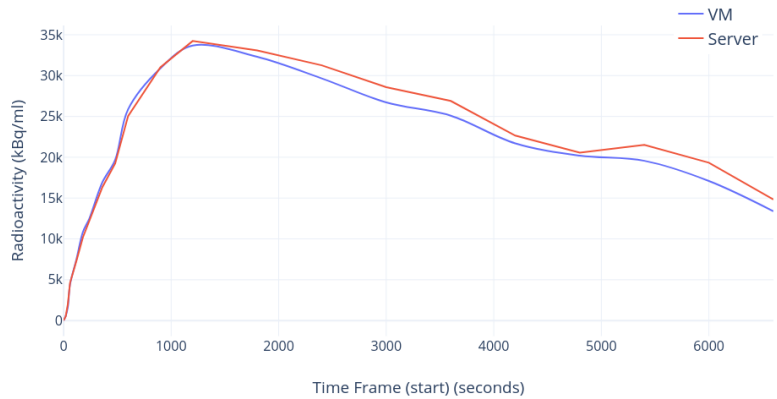


Figure 18: Time activity curve for the Medial Orbitofrontal Cortex across two different computational environments (VM and server), highlighting the uptake of radioactivity within this region as a function of time

Why do these differences arise?

- *significant bias* across *brain regions* and variance in the bias across *subjects*
- Variations due to difference in the cortical and sub-cortical segmentations obtained.
- *Difference in the operating system*, *library and system call interception* and floating point arithmetic as previous studies show

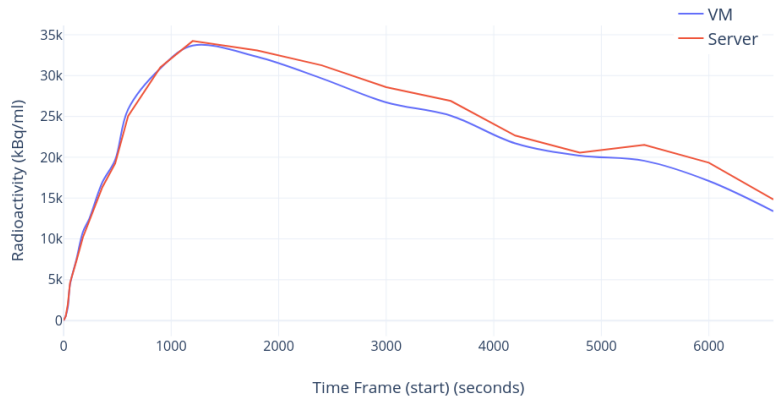
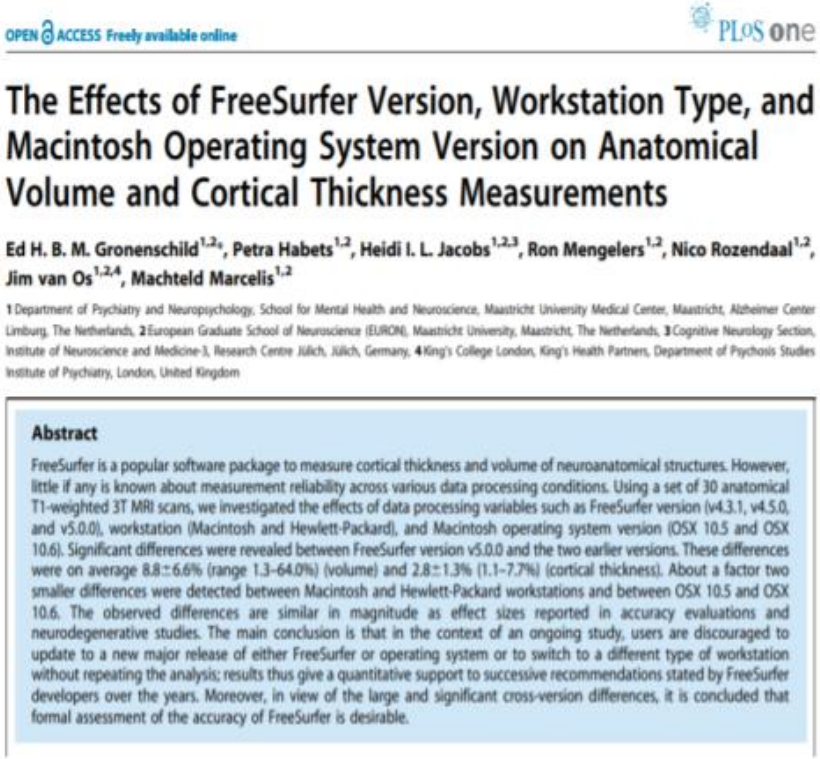


Figure 18: Time activity curve for the Medial Orbitofrontal Cortex across two different computational environments (VM and server), highlighting the uptake of radioactivity within this region as a function of time

Why do these differences arise?



OPEN ACCESS Freely available online

PLoS one

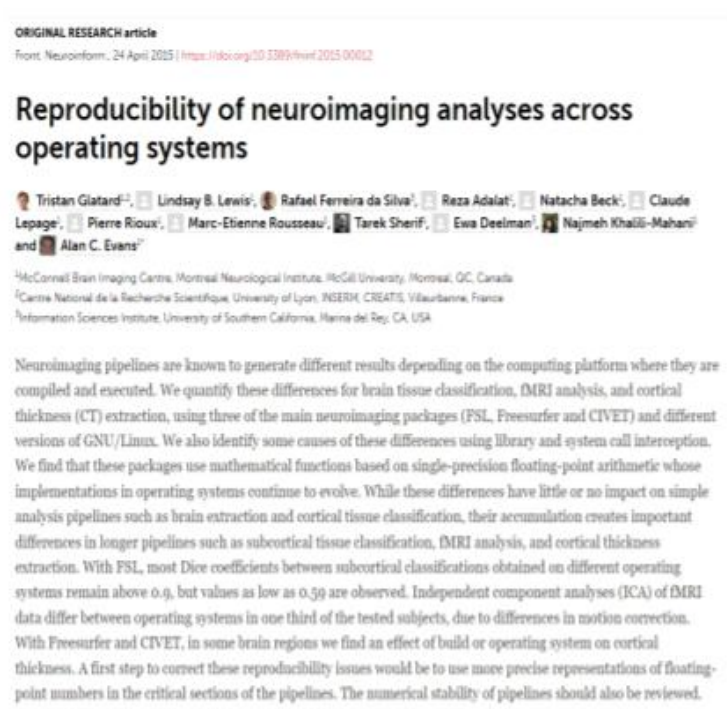
The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements

Ed H. B. M. Gronenschild[1,2]*, Petra Habets[1,2], Heidi I. L. Jacobs[1,2,3], Ron Mengelers[1,2], Nico Rozendaal[1,2], Jim van Os[1,2,4], Machteld Marcelis[1,2]

1 Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Center, Maastricht, Alzheimer Center Limburg, The Netherlands, 2 European Graduate School of Neuroscience (EURON), Maastricht University, Maastricht, The Netherlands, 3 Cognitive Neurology Section, Institute of Neuroscience and Medicine-3, Research Centre Jülich, Jülich, Germany, 4 King's College London, King's Health Partners, Department of Psychosis Studies Institute of Psychiatry, London, United Kingdom

Abstract

FreeSurfer is a popular software package to measure cortical thickness and volume of neuroanatomical structures. However, little if any is known about measurement reliability across various data processing conditions. Using a set of 30 anatomical T1-weighted 3T MRI scans, we investigated the effects of data processing variables such as FreeSurfer version (v4.3.1, v4.5.0, and v5.0.0), workstation (Macintosh and Hewlett-Packard), and Macintosh operating system version (OSX 10.5 and OSX 10.6). Significant differences were revealed between FreeSurfer version v5.0.0 and the two earlier versions. These differences were on average 8.8±6.6% (range 1.3–64.0%) (volume) and 2.8±1.3% (1.1–7.7%) (cortical thickness). About a factor two smaller differences were detected between Macintosh and Hewlett-Packard workstations and between OSX 10.5 and OSX 10.6. The observed differences are similar in magnitude as effect sizes reported in accuracy evaluations and neurodegenerative studies. The main conclusion is that in the context of an ongoing study, users are discouraged to update to a new major release of either FreeSurfer or operating system or to switch to a different type of workstation without repeating the analysis; results thus give a quantitative support to successive recommendations stated by FreeSurfer developers over the years. Moreover, in view of the large and significant cross-version differences, it is concluded that formal assessment of the accuracy of FreeSurfer is desirable.



ORIGINAL RESEARCH article

Front. Neuroinform., 24 April 2015 | https://doi.org/10.3389/fninf.2015.00012

Reproducibility of neuroimaging analyses across operating systems

Tristan Glatard[1,2], Lindsay B. Lewis[1], Rafael Ferreira da Silva[3], Reza Adalat[1], Natacha Beck[1], Claude Lepage[1], Pierre Rioux[1], Marc-Etienne Rousseau[1], Tarek Sherif[1], Ewa Deelman[3], Najmeh Khalili-Mahani[1] and Alan C. Evans[1]*

[1]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada
[2]Centre National de la Recherche Scientifique, University of Lyon, INSERM, CREATIS, Villeurbanne, France
[3]Information Sciences Institute, University of Southern California, Marina del Rey, CA, USA

Neuroimaging pipelines are known to generate different results depending on the computing platform where they are compiled and executed. We quantify these differences for brain tissue classification, fMRI analysis, and cortical thickness (CT) extraction, using three of the main neuroimaging packages (FSL, Freesurfer and CIVET) and different versions of GNU/Linux. We also identify some causes of these differences using library and system call interception. We find that these packages use mathematical functions based on single-precision floating-point arithmetic whose implementations in operating systems continue to evolve. While these differences have little or no impact on simple analysis pipelines such as brain extraction and cortical tissue classification, their accumulation creates important differences in longer pipelines such as subcortical tissue classification, fMRI analysis, and cortical thickness extraction. With FSL, most Dice coefficients between subcortical classifications obtained on different operating systems remain above 0.9, but values as low as 0.59 are observed. Independent component analyses (ICA) of fMRI data differ between operating systems in one third of the tested subjects, due to differences in motion correction. With Freesurfer and CIVET, in some brain regions we find an effect of build or operating system on cortical thickness. A first step to correct these reproducibility issues would be to use more precise representations of floating-point numbers in the critical sections of the pipelines. The numerical stability of pipelines should also be reviewed.

Is the pipeline robust and computationally reproducible?

- Since the pipeline completed running successfully on different datasets, the developed pipeline *seems to be* **robust**.

Is the pipeline robust and computationally reproducible?

- Since the pipeline completed running successfully on different datasets, the developed pipeline *seems to be* **robust**.
- The results can be *replicated* given that the computational environment is consistent in terms of the *software packages*, *operating system* and *configuration parameters* used.

Is the pipeline robust and computationally reproducible?

- Since the pipeline completed running successfully on different datasets, the developed pipeline *seems to be* **robust**.
- The results can be *replicated* given that the computational environment is consistent in terms of the *software packages*, *operating system* and *configuration parameters* used.
- The details of the preprocessing including the inputs and outputs to the various preprocessing steps are captured as a computation graph making the methodology transparent and **reproducible**.

Summary of strengths and weaknesses

× Limited Testing across environments and subjects.

Summary of strengths and weaknesses

- × Limited Testing across environments and subjects.
- × No flexibility in terms of including/excluding preprocessing steps in PETPipeline.

Summary of strengths and weaknesses

- × Limited Testing across environments and subjects.
- × No flexibility in terms of including/excluding preprocessing steps in PETPipeline.
- × No flexibility in terms of software tools for preprocessing steps in PETPipeline.

Summary of strengths and weaknesses

- × Limited Testing across environments and subjects.
- × No flexibility in terms of including/excluding preprocessing steps in PETPipeline.
- × No flexibility in terms of software tools for preprocessing steps in PETPipeline.
- ✓ Flexibility offered in terms of changing parameters of individual steps

Summary of strengths and weaknesses

- × Limited Testing across environments and subjects.
- × No flexibility in terms of including/excluding preprocessing steps in PETPipeline.
- × No flexibility in terms of software tools for preprocessing steps in PETPipeline.
- ✓ Flexibility offered in terms of changing parameters of individual steps
- ✓ Code and documentation available thereby making PETPipeline reproducible

Summary of strengths and weaknesses

- × Limited Testing across environments and subjects.
- × No flexibility in terms of including/excluding preprocessing steps in PETPipeline.
- × No flexibility in terms of software tools for preprocessing steps in PETPipeline.
- ✓ Flexibility offered in terms of changing parameters of individual steps
- ✓ Code and documentation available thereby making PETPipeline reproducible
- ✓ Documentation available for making the pipeline fully executable in a docker container.

Conclusion

- An automated pipeline was developed for preprocessing of PET images using Nipype, and made available on GitHub.

Conclusion

- An automated pipeline was developed for preprocessing of PET images using Nipype, and made available on GitHub.
- The pipeline ran successfully on all datasets indicating that it is robust to errors.

Conclusion

- An automated pipeline was developed for preprocessing of PET images using Nipype, and made available on GitHub.
- The pipeline ran successfully on all datasets indicating that it is robust to errors.
- Using the same computational resources such as OS, software packages, versions and configurations across different computational environments allows us to fully replicate the results indicating that the pipeline is reproducible.

Conclusion & Future Work

- Testing on computational environments with different computational resources showed variations in the resulting BP estimates across brain regions and subjects.

Conclusion & Future Work

- Testing on computational environments with different computational resources showed variations in the resulting BP estimates across brain regions and subjects.
- These may be attributed to a difference in cortical and sub-cortical segmentations obtained by FreeSurfer when run on different computational environments.

Conclusion & Future Work

- Testing on computational environments with different computational resources showed variations in the resulting BP estimates across brain regions and subjects.
- These may be attributed to a difference in cortical and sub-cortical segmentations obtained by FreeSurfer when run on different computational environments.
- More thorough testing would help in further investigating these differences which can aid in further understanding the cause.

References I

[1] Melanie Ganz-Benjaminsen and Martin Nørgaard. “”[11C]DASB PET Cimbi database example””. In: (2021). DOI: 10.18112/openneuro.ds001420.v1.1.0.

[2] Melanie Ganz-Benjaminsen and Martin Nørgaard. “”[11C]SB207145 PET Cimbi database example””. In: (2021). DOI: 10.18112/openneuro.ds001421.v1.2.1.

[3] Martin Nørgaard et al. “Optimization of preprocessing strategies in Positron Emission Tomography (PET) neuroimaging: A [11C]DASB PET study”. In: *NeuroImage* 199 (2019), pp. 466–479.

[4] R Todd Ogden et al. “In vivo quantification of serotonin transporters using [(11)C]DASB and positron emission tomography in humans: modeling considerations”. en. In: *J. Cereb. Blood Flow Metab.* 27.1 (Jan. 2007), pp. 205–217.

Thank you!

Questions?

Additional Slides I

Statistical tests

- One Way ANOVA using computational environments as repeated measures for 2 brain regions. ANOVA (Analysis of Variance) is a statistical test used to analyze the difference between the means of more than two groups
- Simple Linear Regression to find the best-fit slope and intercept for the various subjects and sessions across different computational environments. Comparison of Regression slopes using p values to evaluate the differences.

Additional Slides II

Choice of programming language, and library (Nipype) for implementation

- Python: open source language, community support. Implementation of various neuroimaging tools such as PyBids, Nibabel, Nilearn available for neuroimaging analysis.
- Nipype: Interfaces to various command-line based neuroimaging software such as FSL, AFNI, Freesurfer available. Enables interaction between these tools in a single workflow, combine multiple workflows into a single workflow. Allows for parallel processing across multiple subjects in a dataset.

Additional Slides III

Why BIDS?

- All data may not be necessarily available in the same format. Different datasets can be in different formats such as Dicom, Analyze etc.
- All the datasets need to be organized in a single format so that standardized neuroimaging pipelines can be run on them. BIDS provides a way to organize and structure the complex neuroimaging data coming from different neuroimaging modalities. Widely adopted by the neuroimaging community, a variety of tools have been developed around it to promote data sharing, validation, processing and analysis.
- BIDS also lists a specification for the derivatives of the various image processing techniques applied on the datasets.
- Moreover, using a standardized structure promotes reproducibility as it allows researchers to share each other's experimental results, and promotes the development of software that can process neuroimaging data following the BIDS structure.