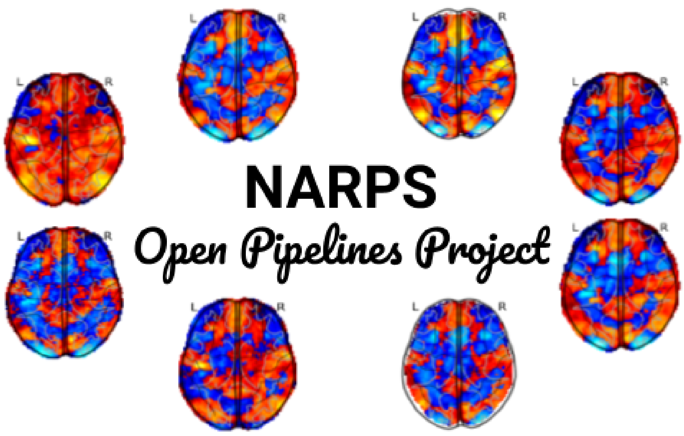
Reproducing brain imaging pipelines to study analytical variability

Supervision : **Boris Clénet**, Research engineer at Inria, Empenn team (boris.clenet@irisa.fr)
Camille Maumet, Research scientist at Inria, Empenn team (camille.maumet@inria.fr)
Location : Empen team, Inria Rennes / IRISA, Campus de Beaulieu, 35042 Rennes, France
Duration : 5 to 6 months
Research domains: Reproducibility in Computer Science, Statistics, Neuroinformatics.
Keywords: Analytical variability, data processing, open science, reproducibility, brain

Context

10 years ago, neuroinformatics – i.e. the equivalent of the field of bioinformatics but for brain imaging – faced a reproducibility crisis which put into question the validity of published findings (Button et al., 2013; Poldrack et al., 2017). Since then many progress have been made to improve the reliability of neuroimaging findings, including open science practices such as the sharing of open data.

Yet, throughout the years, the different tools and approaches available to analyse brain imaging data have multiplied. Each approach provides its own version of the results and overall pipeline multiplicity leads to a very large space of possible results leaving practitioners at a loss to find the right answer to their research question. Until recently, this analytical variability – i.e. the variability induced by different pipelines on the results – has typically been ignored, effectively considering that it was negligible compared to other sources of variability. But In recent years, increasing evidence has shown that the exact choice of pipeline has a non-negligible impact on research findings. Analytical variability affects many fields in science (Hoffmann et al., 2021).



For brain imaging research, analytical variability has been shown to manifest in different contexts (Carp, 2012). A particularly striking finding here observed that results varied depending on whether they were obtained using a CentOS or a Fedora operating system (Glatard et al., 2015). Other authors have shown that the version of domain-specific software also had a non-negligible impact on research findings. In (Bowring et al., 2021, 2022), we examined variations across the three main neuroimaging software packages used in functional Magnetic Resonance Imaging (fMRI).

In 2020 the NARPS study (Botvinik-Nezer et al., 2020) aimed at highlighting the variability in the findings obtained by the re-analysis of the same neuroimaging dataset by different expert teams. Each team chose a different pipeline and – what is more worrying – those differences in pipelines led to contradictory findings. This creates an urgency to find solutions to guide practitioners in the multiverse of pipelines in order to identify results that are both valid and generalizable.

In 2021, the Empenn team started to develop as a tool to help the research community explore analytical variability in the context of NARPS. This led to the NARPS Open Pipelines project, an open-source and contributive codebase reproducing the 70 pipelines of the NARPS study (https://github.com/Inria-Empenn/narps_open_pipelines).

The goal of the internship **will be to reproduce fMRI experiments from the NARPS study, contribute to NARPS open pipelines and study analytical variability in brain imaging.**

Main activities

Initially, you will need to familiarise yourself with the field of functional brain imaging and in particular with the processing chains used in fMRI (preprocessing, statistical analysis, etc.). In addition, you will carry out a literature review on the vibration of results.

The first objective of this work will be to reproduce the work of Chen and colleagues (Chen et. al. 2022) which focused on using FSL (one of the three most popular neuroimaging software packages) within Nipype (the workflow engine used in NARPS open pipelines). To do this, you will have to program an fMRI processing chain on real data (Python), document your efforts and check that the results obtained are consistent between FSL alone and FSL in Nipype.

In a second stage, you will be able to contribute to NARPS Open Pipelines, specifically the FSL-based pipelines. Finally, you will work on the NARPS open pipeline codebase to study analytical variability in fMRI (see for example (Germani & Maumet, 2022)).

Requirements

We look for strong candidates motivated by research topics in brain imaging, in the context of open science. The applicant should present a background in image processing and/or computer science. Basic knowledge in Python language is required.

References

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), Article 7810. https://doi.org/10.1038/s41586-020-2314-9

Bowring, A., Nichols, T. E., & Maumet, C. (2022). Isolating the sources of pipeline-variability in group-level task-fMRI results. *Human Brain Mapping*, *43*(3), 1112–1128. https://doi.org/10.1002/hbm.25713

Bowring, A., Nichols, T., & Maumet, C. (2021). *Identifying sources of software-dependent differences in task fMRI analyses*. 1. https://inserm.hal.science/inserm-03479022

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, 13.

Germani, E., & Maumet, C. (2022). *FMRI data analysis: How does analytical variability vary with sample size?* 1. https://www.hal.inserm.fr/inserm-03642535

Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., Rioux, P., Rousseau, M.-E., Sherif, T., Deelman, E., Khalili-Mahani, N., & Evans, A. C. (2015). Reproducibility of neuroimaging analyses across operating systems. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00012

Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U., & Boulesteix, A.-L. (2021). The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines. *Royal Society Open Science*, *8*(4), rsos.201925, 201925. https://doi.org/10.1098/rsos.201925

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

If you are interested in this internship or have any questions, please do not hesitate to contact the supervisors by email (see top of this document).