



**HAL**  
open science

# Representation learning for more reproducible fMRI data analyses

Elodie Germani, Elisa Fromont, Camille Maumet

► **To cite this version:**

Elodie Germani, Elisa Fromont, Camille Maumet. Representation learning for more reproducible fMRI data analyses. IABM 2023 - Colloque Français d'Intelligence Artificielle en Imagerie Biomédical, Mar 2023, Paris, France. pp.1. hal-04068797

**HAL Id: hal-04068797**

**<https://inria.hal.science/hal-04068797>**

Submitted on 14 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## UNDERSTANDING VARIABILITY

How does analytical variability vary with sample size ?

(Germani et al, 2022)

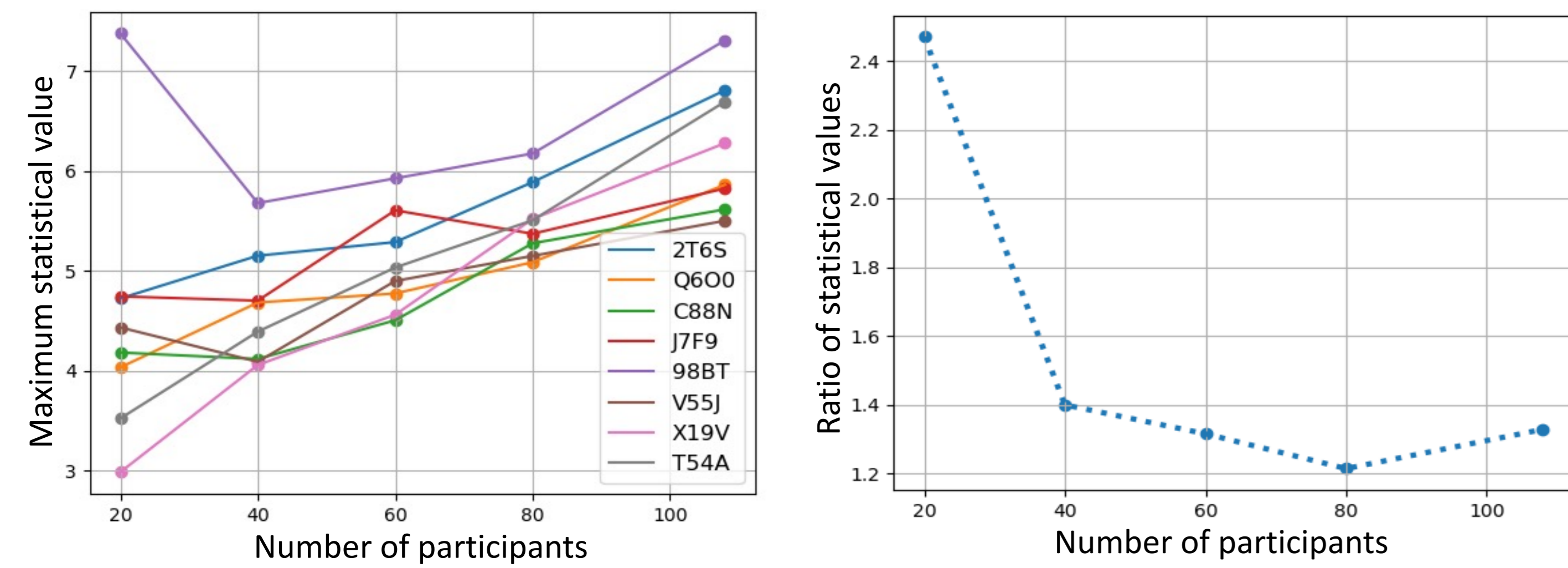


Fig. 1 - Maximum statistical value inside ROI depending on sample size (right) and ratio of largest/smallest maximum statistical value depending on sample sizes (left)

In psychology, (Klau et al., 2020) showed that the vibration of effects decreases and stabilizes as sample size increases.

- ✓ Reproduction of 8 pipelines used in a many-analyst study  
→ NARPS Open Pipelines: [github.com/Inria-Empenn/narps\\_open\\_pipelines](https://github.com/Inria-Empenn/narps_open_pipelines)
- ✓ Selection of subsets with different numbers of participants;
- ✓ Replication of the results with the subsets.

→ With growing sample sizes, maximum statistical values seemed to converge  
→ Confirmed by the stabilization and convergence to 1.3 of the ratio between the largest and the smallest maximum statistical values among teams with sample size.  
→ But even for large sample sizes, analytical variability remains non neglectable.

Code available at [gitlab.inria.fr/egermani/analytic\\_variability\\_fmri/](https://gitlab.inria.fr/egermani/analytic_variability_fmri/)

### Exploring variability patterns in the analytical space

(Germani et al, 2023b)

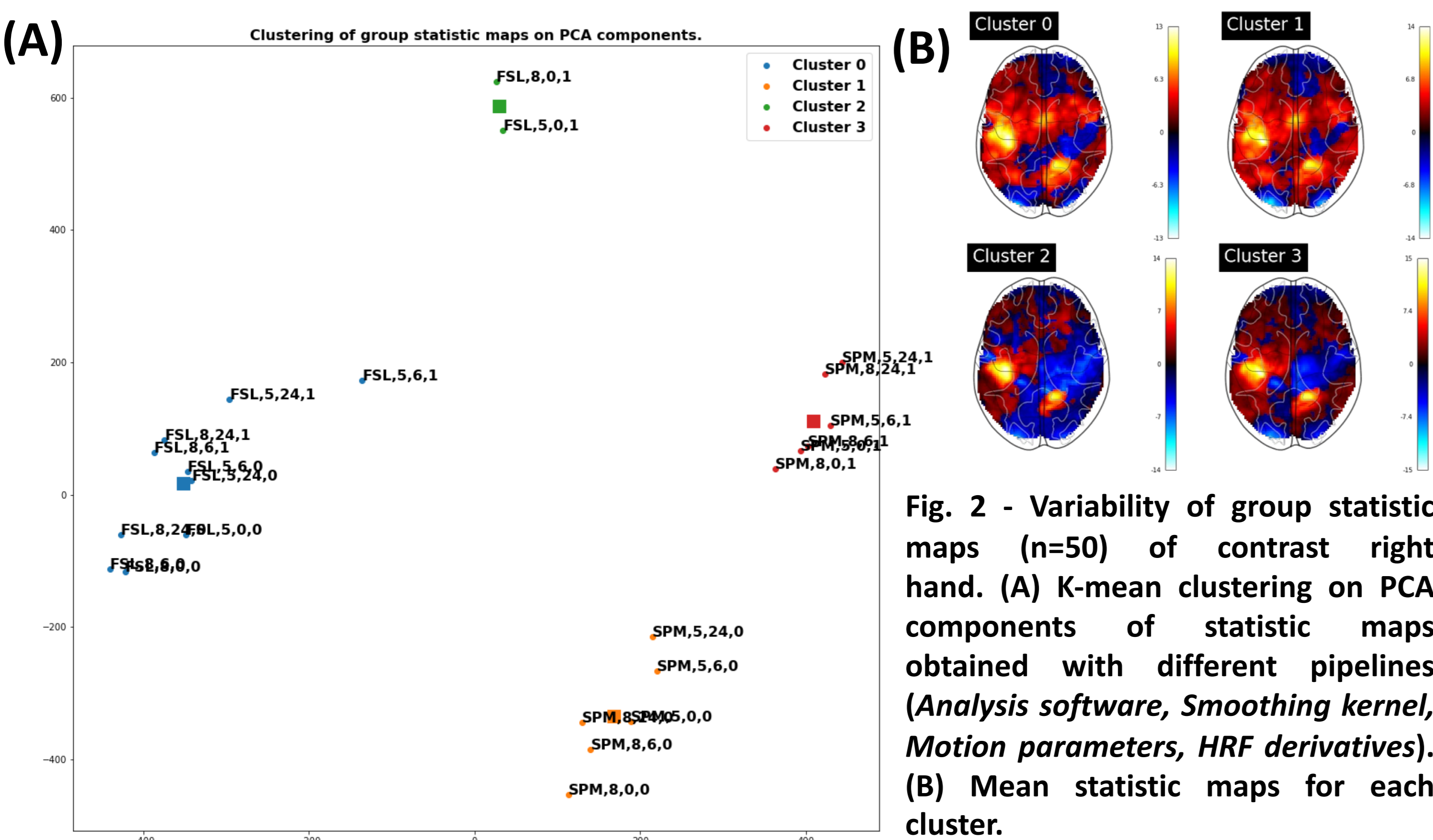


Fig. 2 - Variability of group statistic maps (n=50) of contrast right hand. (A) K-mean clustering on PCA components of statistic maps obtained with different pipelines (Analysis software, Smoothing kernel, Motion parameters, HRF derivatives). (B) Mean statistic maps for each cluster.

The fMRI pipeline-space is especially large and challenging to investigate due the pipeline's sensitivity to particular properties of the data under study.

- ✓ Use of 24 different fMRI pipelines with varying parameters (software, smoothing kernel, etc.) on data from the Human Connectome Project (n=1080 participants)
- ✓ Dimension reduction using PCA or AutoEncoders and clustering methods

→ Pipelines have different behaviors depending on contrasts  
→ Some pipelines give closer results than others, especially those sharing specific parameters.

Code available at [gitlab.inria.fr/egermani/hcp\\_pipelines/](https://gitlab.inria.fr/egermani/hcp_pipelines/)

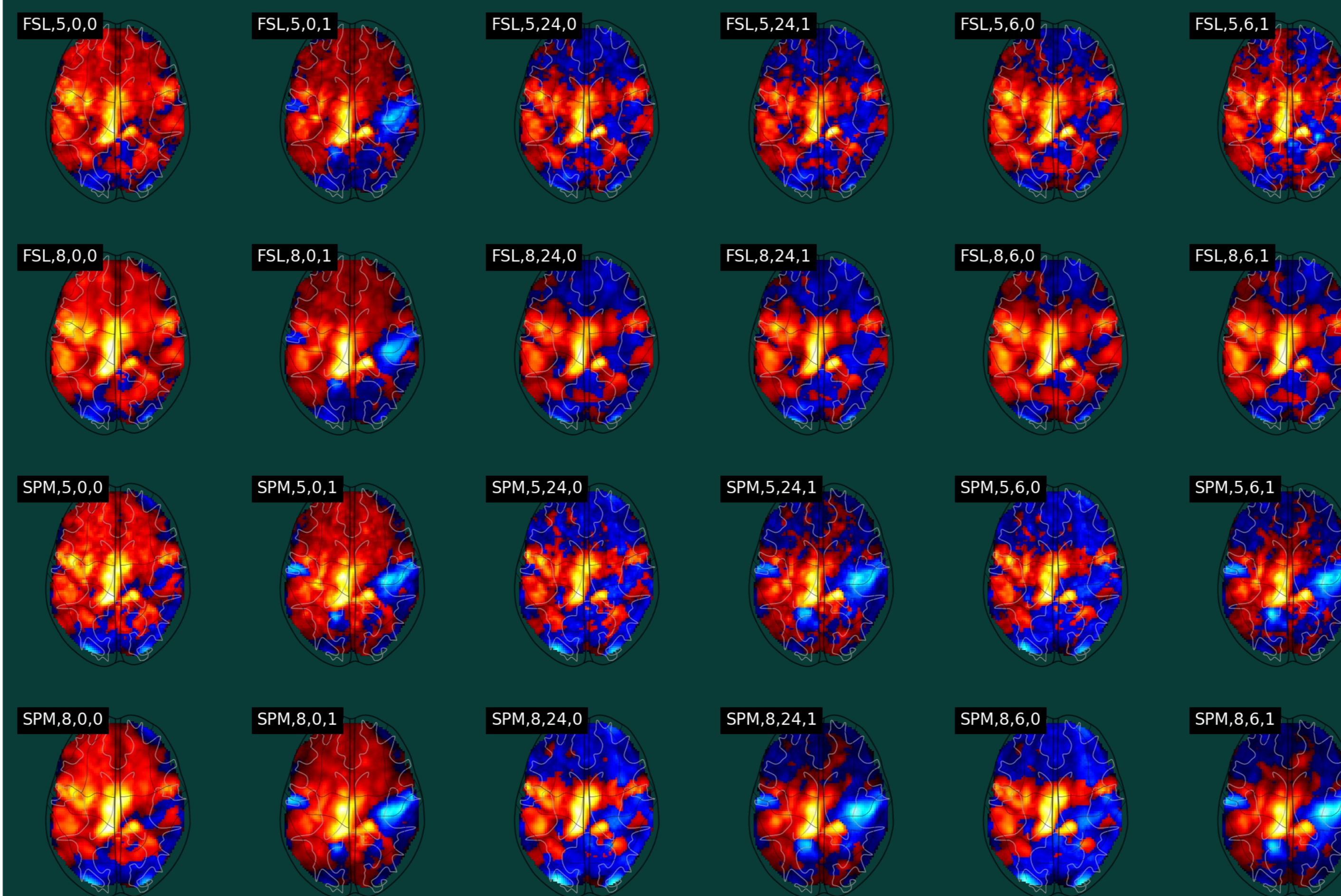
# Representation learning for more reproducible fMRI data analyses

Elodie Germani<sup>1</sup>, Elisa Fromont<sup>2\*</sup>, Camille Maumet<sup>1\*</sup>

<sup>1</sup> Univ Rennes, Inria, CNRS, Inserm, Rennes, France

<sup>2</sup> Univ Rennes, IUF, Inria, CNRS, IRISA, Rennes, France

\* Joint senior authorship



## SAME RAW DATA, DIFFERENT ANALYTIC PIPELINES

- Neuroimaging workflows are highly flexible with many possible choices at each step of an analysis (Carp 2012).
- However, different analytical choices can cause variation in the results (Botvinik-Nezer et al., 2020).
- This puts into question the validity of published results and calls for dedicated methodologies to handle this new level of variability (Dafflon et al. 2020)

→ A better understanding of the relationships between results coming from different pipelines is essential in order to develop new methods to tackle this variability.

## METHODS TO TACKLE THIS VARIABILITY

Benefitting from data variability to build more generalizable brain decoding models (Germani et al, 2023a)

Training effective Deep Learning models on fMRI data comes with challenges, e.g. lack of training data and large number of sources of variability in the data.

- ✓ Use of the NeuroVault (Gorgolewski et al., 2015) database: N=28,532 maps with unconstrained annotations.
- ✓ Pre-training of an unsupervised model (AutoEncoder) using NeuroVault and fine-tuning for classification tasks on smaller datasets.

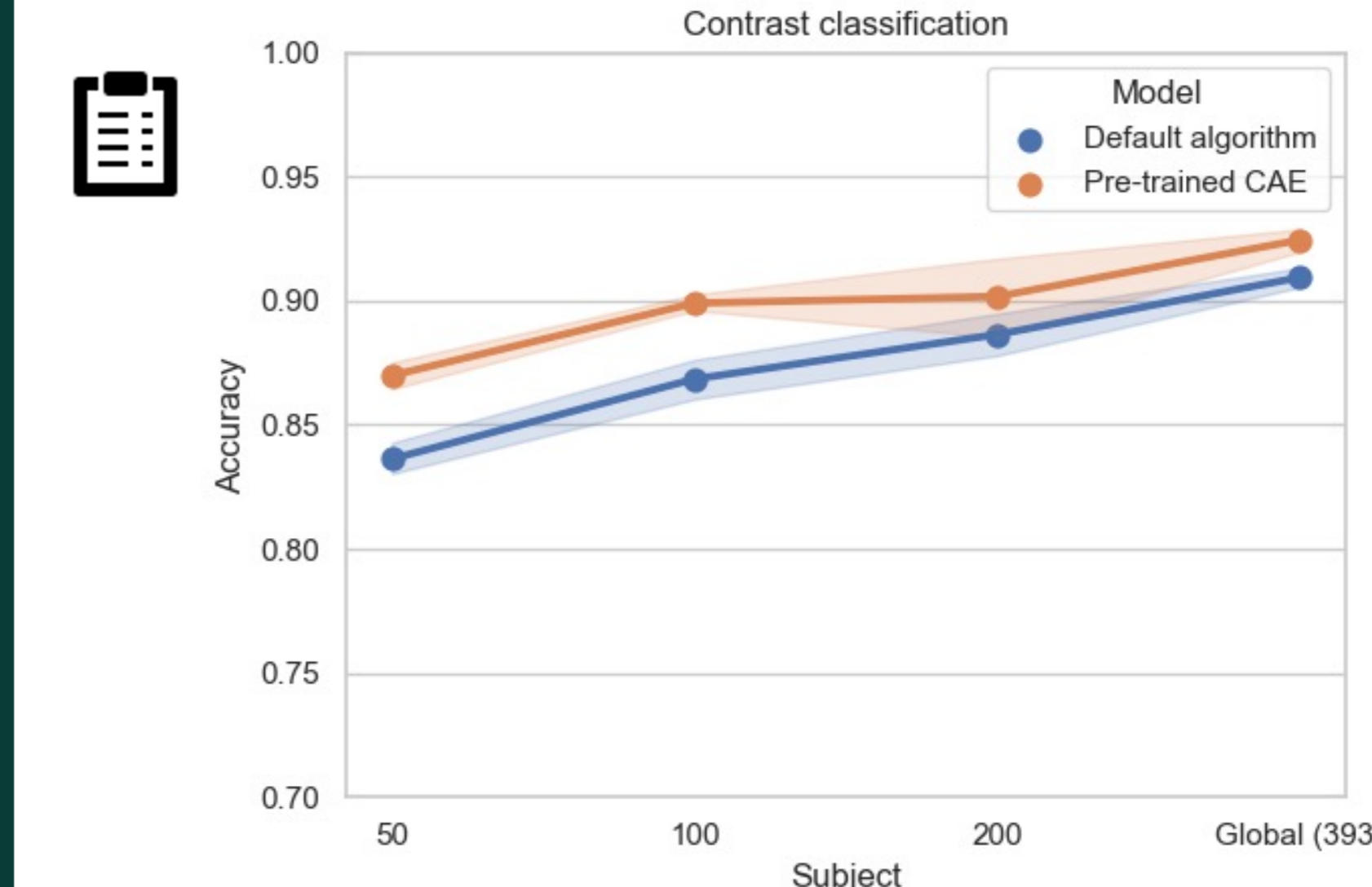


Fig. 3 – Mean accuracies for models initialized with default algorithm (blue) and pre-trained CAE (orange)

→ Better performance on small datasets and difficult classification tasks

→ Improved generalizability due to the ability of the pre-trained to focus on more general features and not on individual patterns of variability.

Code available at [gitlab.inria.fr/egermani/self\\_taught\\_decoding/](https://gitlab.inria.fr/egermani/self_taught_decoding/)

### Transferring pipelines patterns between contrast maps

Re-using fMRI maps processed with different pipelines can be difficult and augment the risk of having false positive findings (Rolland et al., 2022)

- ✓ We hypothesize that different pipelines can be seen as different styles applied to an fMRI map.
- ✓ Use of style transfer methods (Isola et al, 2018) to convert contrast maps between pipelines.

→ Some patterns of a pipeline can easily be transferred between contrast maps (e.g. smoothing)

→ Other parameters may display different patterns in contrast maps depending on specific properties of the data, making style transfer difficult to learn

Code available at [gitlab.inria.fr/egermani/pipeline\\_transition/](https://gitlab.inria.fr/egermani/pipeline_transition/)

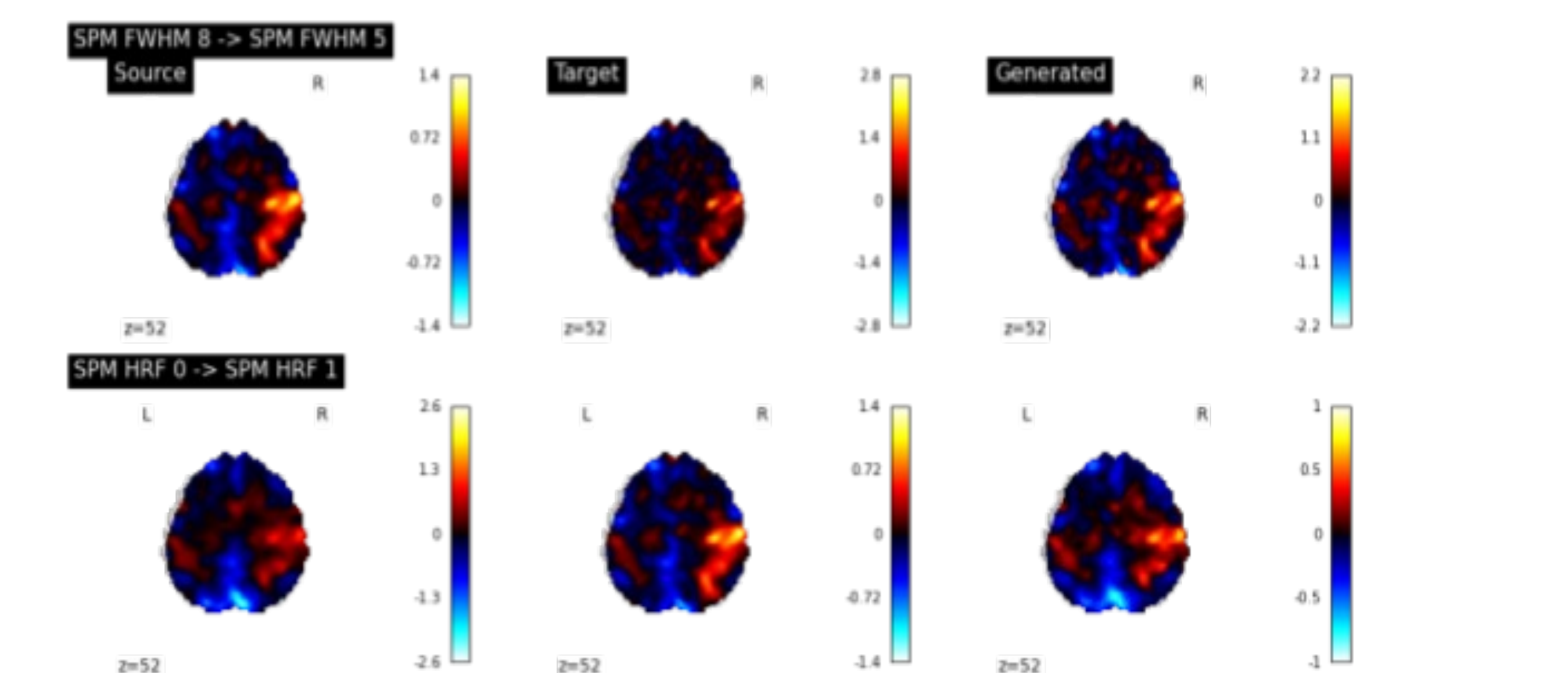


Fig. 4 – Comparison of contrast maps obtained with a source pipeline, a target pipeline and the map generated using style transfer algorithm.

## References

Botvinik-Nezer, R. et al. 2020. « Variability in the analysis of a single neuroimaging dataset by many teams ». Nature.  
 Carp, J. 2012. « On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of FMRI Experiments ». Frontiers in Neuroscience.  
 Dafflon, J. et al. 2020. « Neuroimaging: into the Multiverse ». bioRxiv.  
 Germani, E. et al. 2022. « fMRI data analysis: How does analytical variability vary with sample size? ». OHBM 2022 - 28th Annual Meeting of the Organization for Human Brain Mapping.  
 Germani, E. et al. 2023a. « On the benefits of self-taught learning for brain decoding. » arXiv. (In minor revision at GigaScience)  
 Germani, E. et al. 2023b. « Exploring variability patterns in the task-fMRI analytical space » (submitted at OHBM 2023)  
 Gorgolewski, K.J. et al. 2015. « NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain ». Frontiers in Neuroinformatics.  
 Klau, S. et al. 2020. « Comparing the vibration of effects due to model, data pre-processing and sampling uncertainty on a large data set in personality psychology ». Department of Statistics: Technical Reports.  
 Isola, P. et al. 2018. « Image-to-Image Translation with Conditional Adversarial Networks ». arXiv.  
 Rolland, X. et al. 2022. « Towards efficient fmri data re-use: can we run between-group analyses with datasets processed differently with spm? ». ISBI 2022.