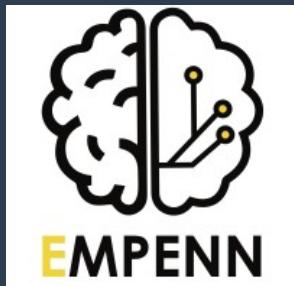


Exploring the variability induced by different configurations in the neuroimaging analytical space

RIPOST Reproducibility Workshop - 2025-10-29



Youenn MEREL JOURDAN
supervised by
Camille MAUMET (Empenn)
Mathieu ACHER (DiverSE)

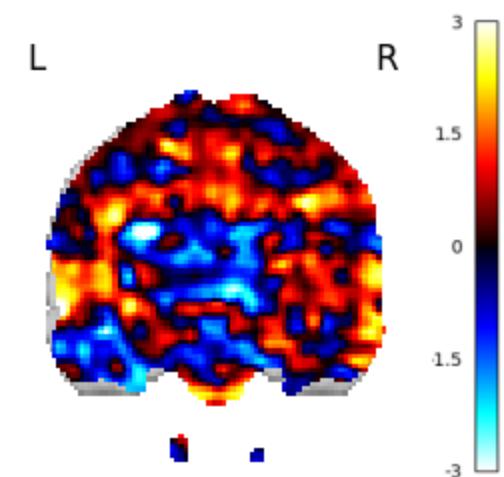


Neuroimaging

- fMRI = functional Magnetic Resonance Imaging
- Study brain activity in subjects performing specific tasks
- Tracks oxygen consumption in the brain across time
- Output 3D matrices = statistic maps



project.inria.fr/neurinfo



Neuroimaging analysis pipelines overview

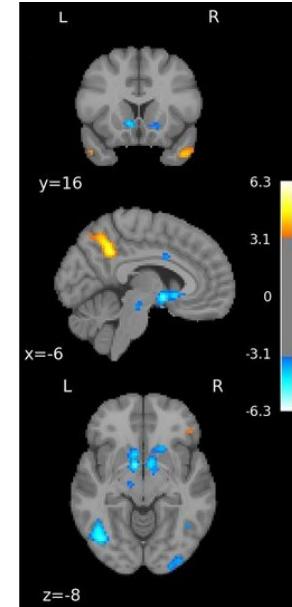
Data are processed with multi-stage software pipelines

Preprocessing

- reduce noise
- improve signal
- standardize data

Subject level & group level analysis

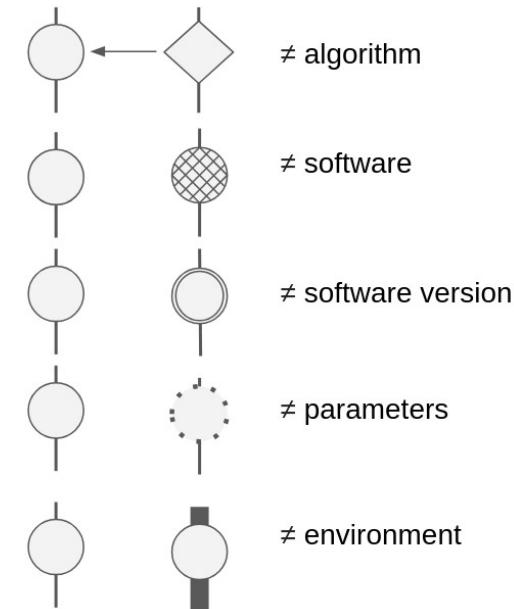
Applies **statistical models** to detect and map activations in the brain



Gallopel-Morvan et al., 2024

Neuroimaging analysis pipelines overview

- A lot of software packages and versions are available
 - AFNI (C, Python, R), FSL (Python), SPM (Matlab), fMRIPrep (NiPype)...
 - Ad-hoc, team specific tools & scripts
- At each step, a lot of parameters, algorithms are available
 - e.g. interpolation algorithm, smoothing kernel, nuisance regressors...
- Processing steps, software, parameters can be combined & ordered in a many ways

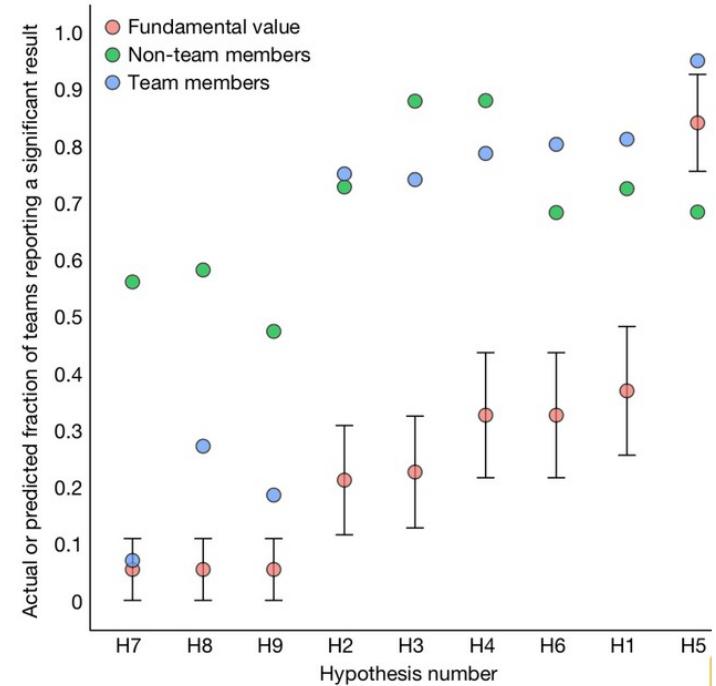


No well supported consensus on pipeline configuration exists yet in the community (no ground truth !)

Neuroimaging pipelines variability

“Variability in the analysis of a single neuroimaging dataset by many teams”
a.k.a. NARPS ([Botvinik-Nezer et al., 2020](#))

- 70 international teams
- 1 fMRI dataset
- Asked to analyze the data using their preferred pipeline
 - Answering same 9 binary research questions
- 70 different pipelines
- A lot of variability in the outcome
 - But with correlated statistical brain maps



Studying neuroimaging pipelines variability

- Results cannot be evaluated against a **directly observable ground truth**
 - **No direct measure** of whole brain activity
- Statistic maps are **complex** and **high-dimensional**
 - as opposed to runtime, number of defects...

How do we evaluate configurations against each other
to study variability?

First contribution

In the Search for Truth: Navigating Variability in Neuroimaging Software Pipelines

Youenn Merel Jourdan, Mathieu Acher, Camille Maumet

29th ACM International Systems and Software Product Lines
Conference (**SPLC 2025**) - A Coruña , Spain, 5th sept. 2025



inria.hal.science/hal-05158426

Research questions

RQ1 (Variability Implementation):

How do we **model & automate** the exploration of **neuroimaging pipeline configurations** using **SPL**?

RQ2 (Accuracy):

How accurate is our **learning** approach in inferring **quantitative performance** against different **proxy ground truths**?

RQ3 (Interpretability):

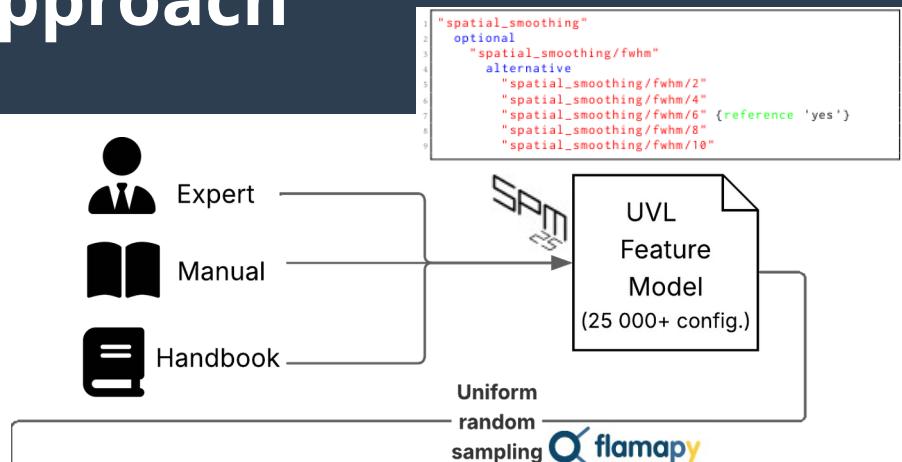
How do we identify **key pipeline features** affecting neuroimaging results?

RQ4 (Cost):

What are the **computational costs** of our approach?

Approach

Modeling



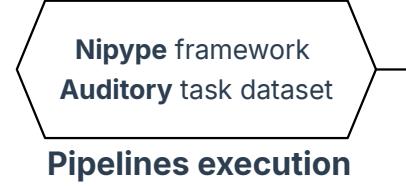
Sampling

slice_timing_correction/ref_slice/middle	signal_modeling/hrf/temporal_dispersion_derivs	...	spatial_smoothing/fwhm	signal_modeling/hrf/canonical
False	False	...	0	True
False	False	...	12	False
...
True	False	...	2	False
True	False	...	6	False

Sample (1000 valid configurations)

Running

```
1 def get_smoothing(self, features: list):
2     node = Node(interface=Smooth(), name="spatial_smoothing")
3     node.inputs.fwhm = 0
4     if f"{name}/fwhm" in features:
5         node.inputs.fwhm = float(self.get_feature_end(f"{name}/fwhm", features))
6     return node
```

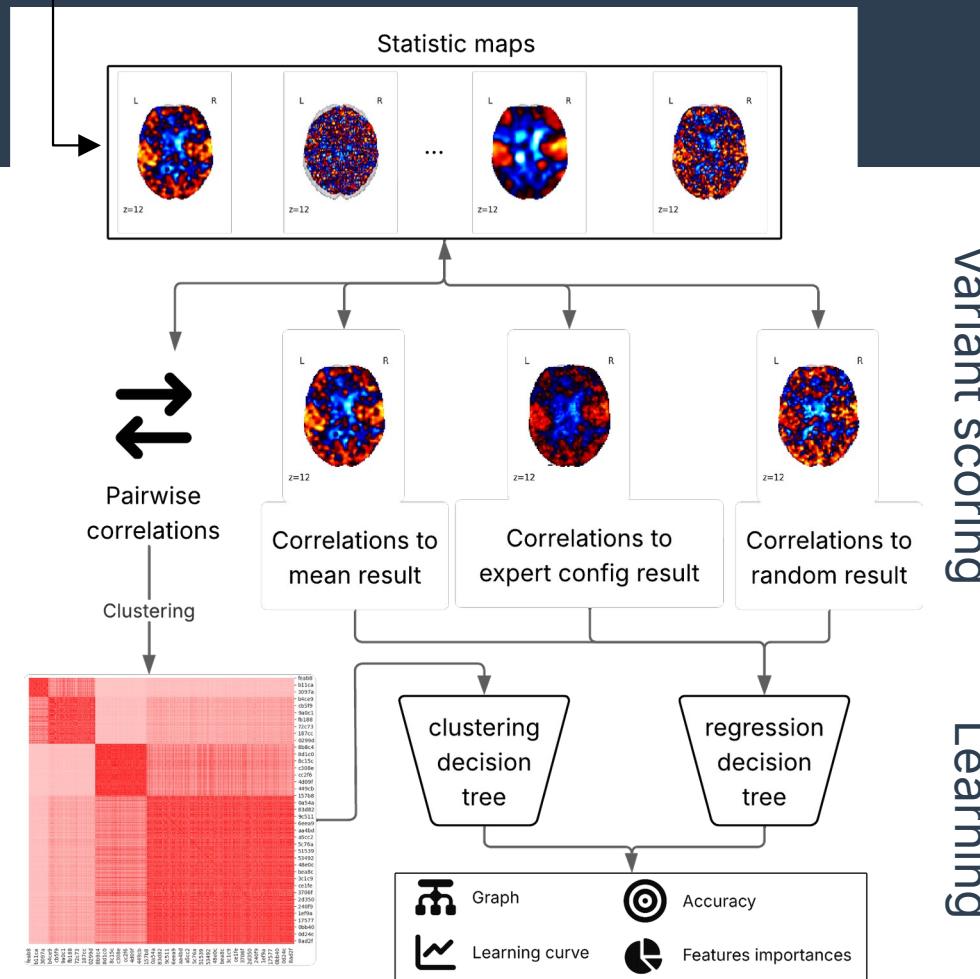


Software Product Line approach

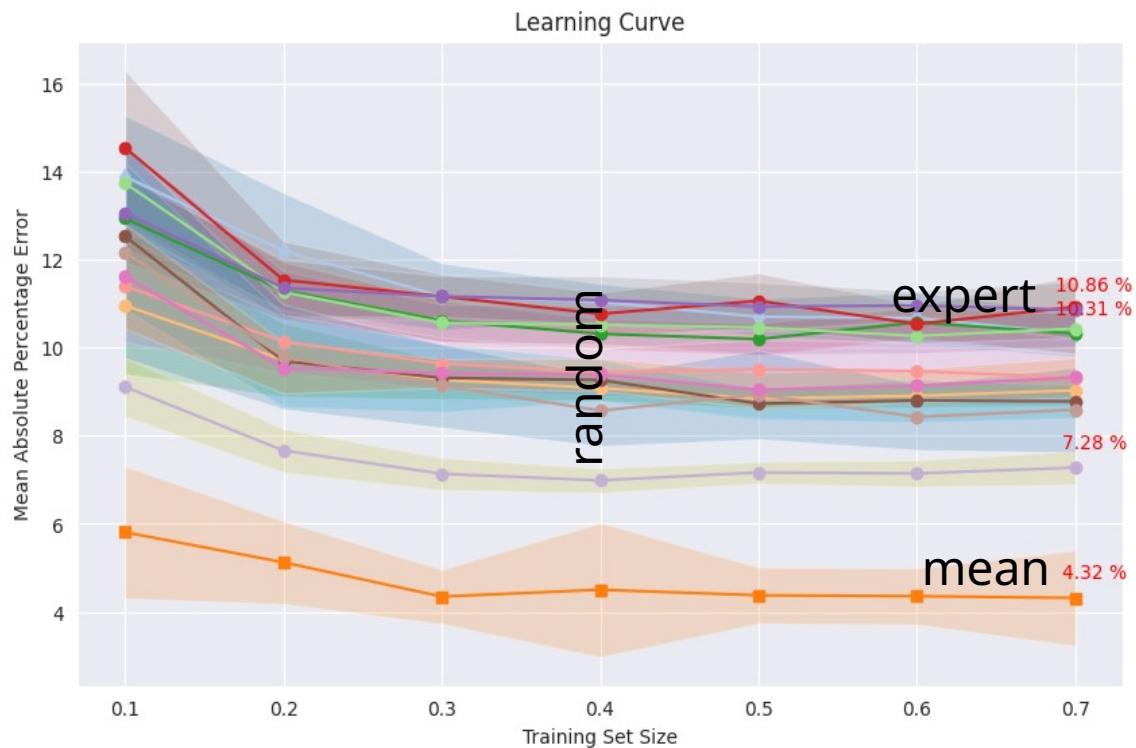
- **Model** our configuration space
- **Sample** from it
- **Build** pipelines
- **Run** them on a task-fMRI dataset

Approach

- Reduce high dimensional statistic maps to a **single score**:
Spearman correlation
- Divide configurations into **4 clusters**
- Try to **predict** correlations, cluster
- **Learn** from it



Main results

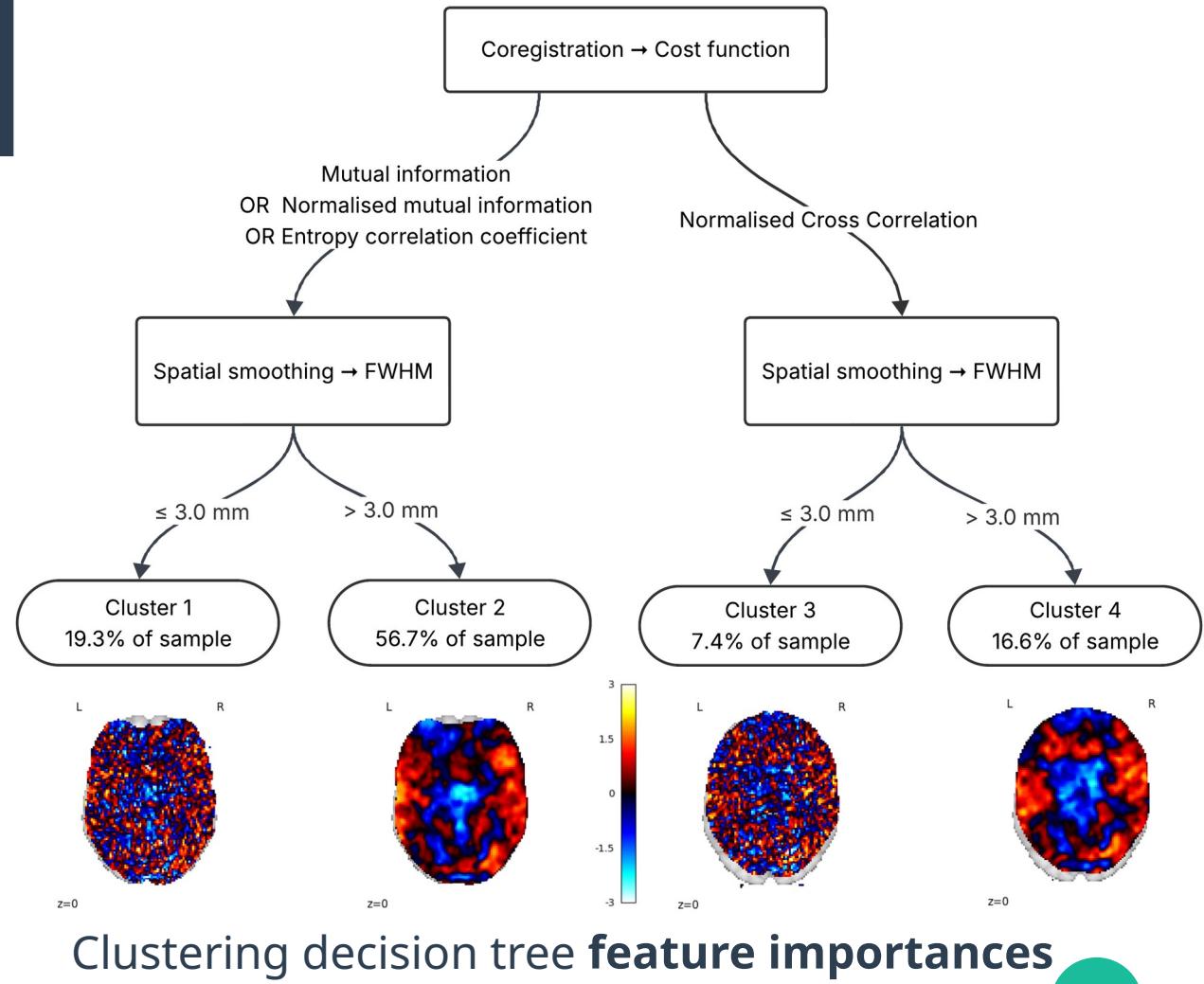


- Predict correlations (mean, expert, random) for any given configuration
- Evaluate relevance of proxy ground truths
- Infer feature importances

Regression decision tree learning curve

Main results

- Predict cluster for any given configuration
- Infer **feature importances**
 - Coregistration (misalignment)
 - Smoothing

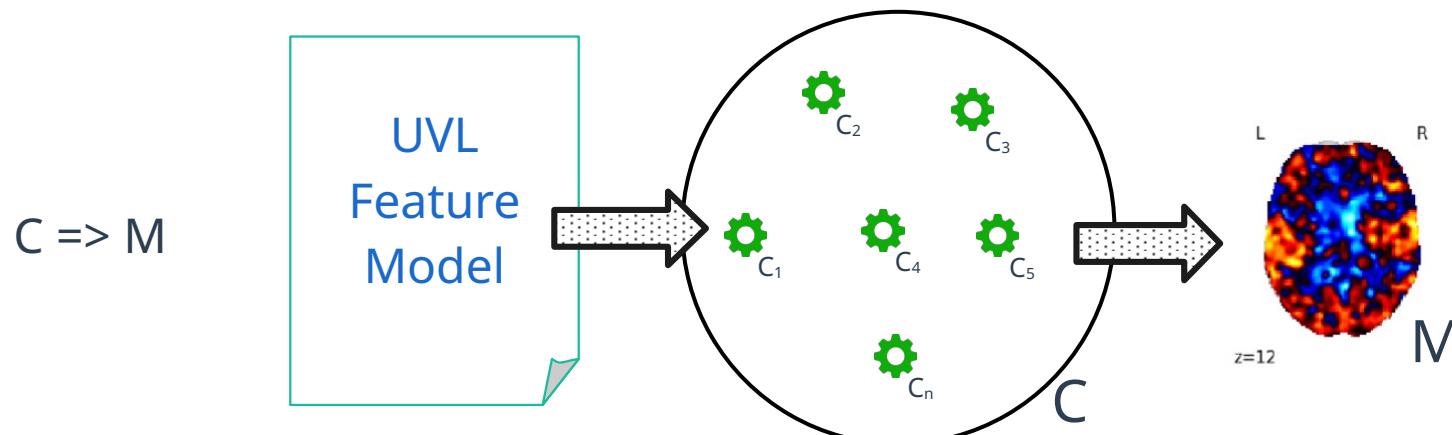


Key take-away

- Our approach to model the neuroimaging pipeline space was successful:
 - 1) Our "**sampling, variants scoring, and learning**" method **accurately generalized** to whole configuration space (RQ2)
 - 2) We **explained** clusters of configuration space with a **few features** (RQ3).
- This was done
 - using an **UVL Feature Model** of a (single-software) neuroimaging configuration space (RQ1)
 - at **manageable computational cost** (RQ4).

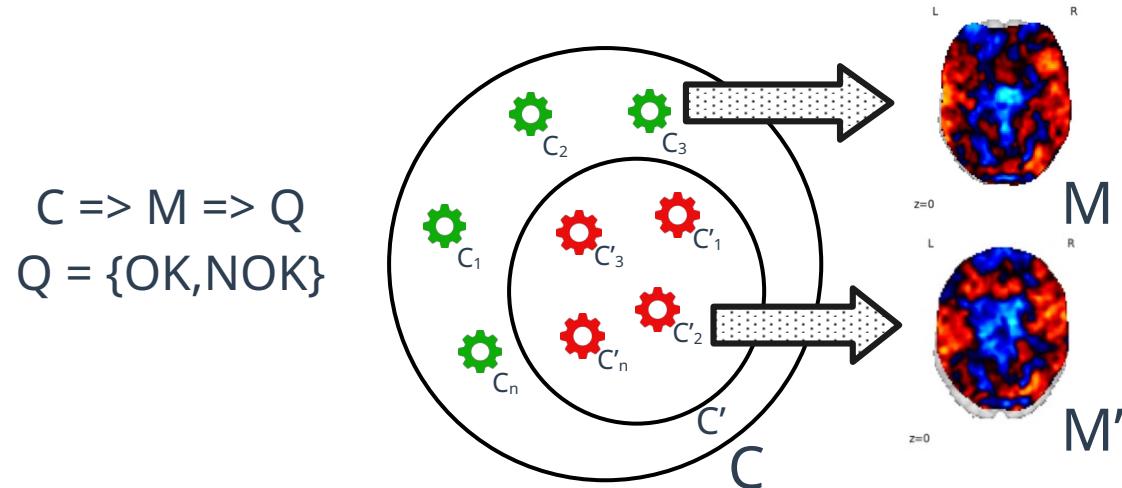
Filtering incorrect configurations

- Automatically detect & exclude incorrect configurations
i.e. configurations that produces a non-acceptable result
 - Work with a more **representative** space
 - Optimize **computational cost**



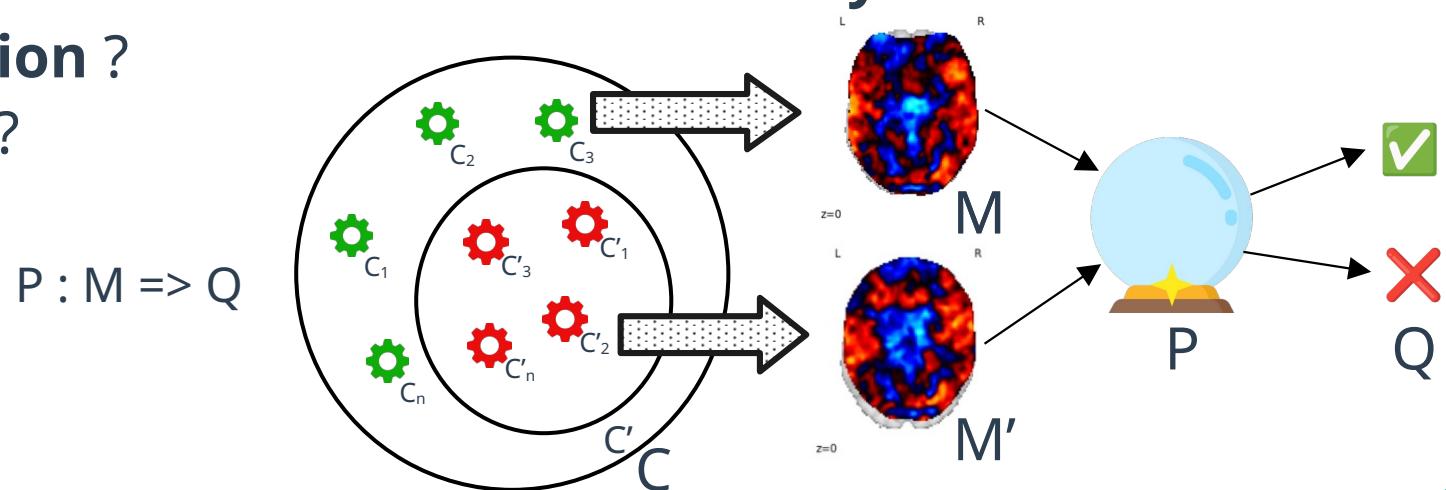
Non acceptable results

- How to **identify** these configurations ?
 - How to **characterise** un-acceptable results ?
 - Misalignment, incomplete brain coverage, whole map activation value too high or too low...



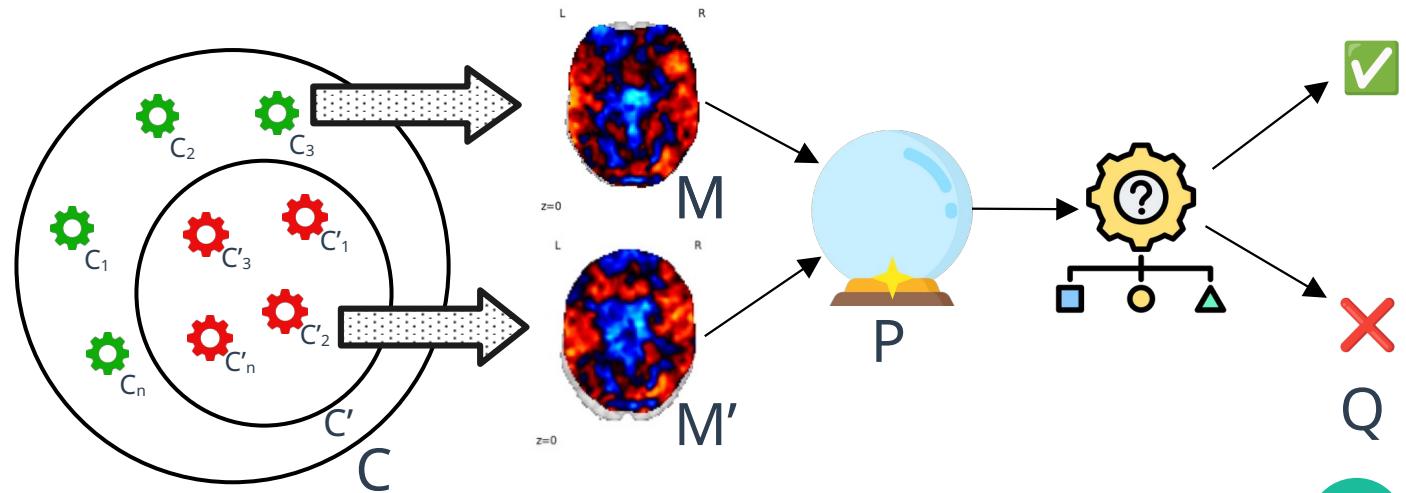
Oracle

- Implement an **oracle** that separates acceptable from un-acceptable results
- How many configuration/results to train for **full space coverage** (including edge cases) ? At what **computational cost** ?
- How to **validate** our oracle ? Evaluate **accuracy** ?
 - **Fault injection** ?
 - **Expert** role ?



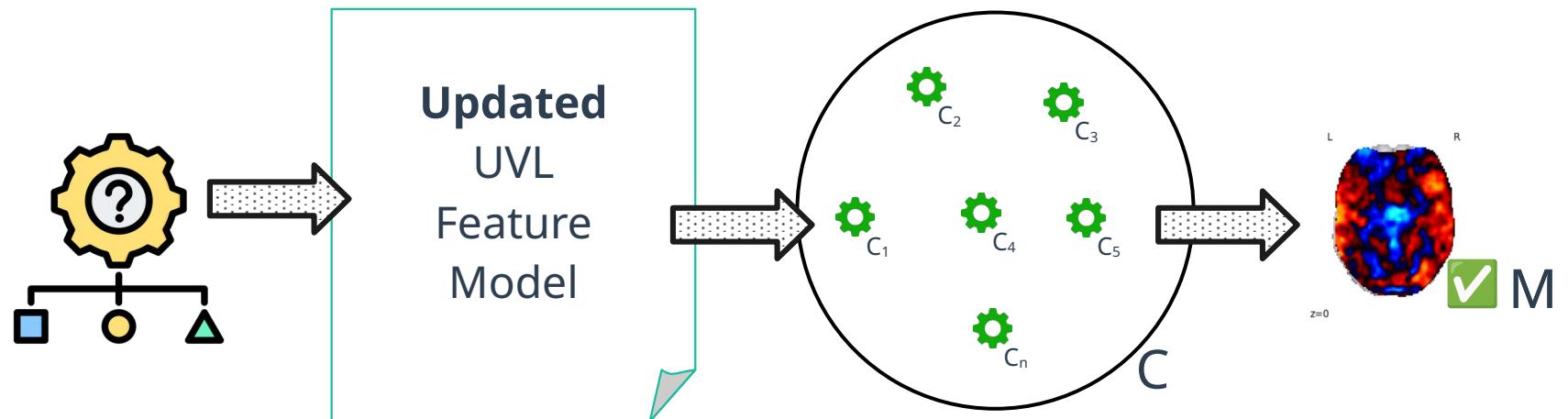
Constraint mining

- Infer **actionnable constraints** from oracle
 - Decision tree (Temple et al., 2016), clustering, associations rules mining, pattern mining
- **Clusterize for dimensionality reduction**
 - Reduce human validation cost ?



Updating Feature Model

- Update FM **constraints** to **reduce** configuration space to **relevant / realistic** configurations
 - Need for **interpretable, generalizable** rules



Data playground

- SPLC'25 dataset : 1 subject, 1 task, 1000 results
- Available open datasets
 - **Human Connectome Project** (Van Essen et al., 2013)
 - **Neurovault** (Gorgolewski et al., 2015)
 - **NARPS** (Botvinik-Nezer et al., 2019)
 - **Narps Open Pipeline** project
(github.com/Inria-Empenn/narps_open_pipelines)

Human Connectome Project

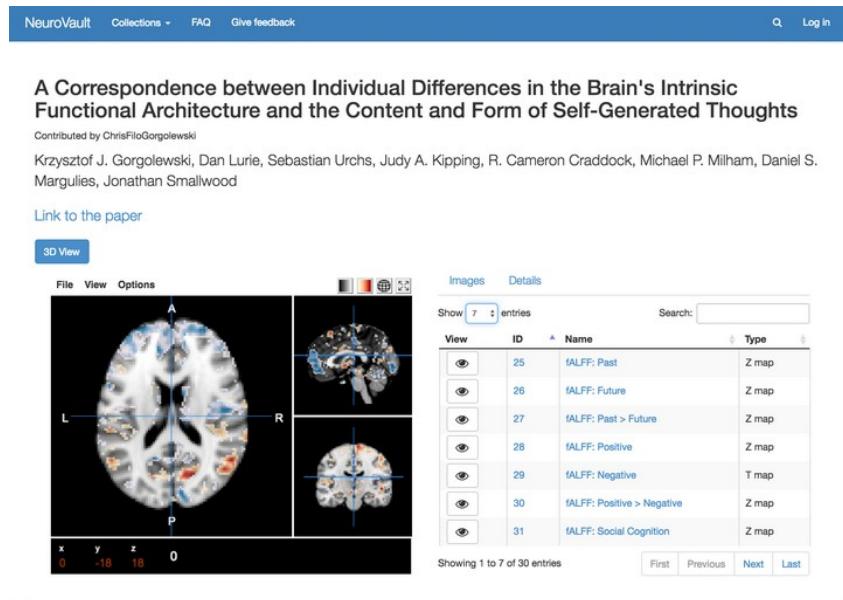
- MRI data from ~1200 subjects
 - **Inter-subject** variability
 - **Group** analysis
- Anatomical + rest fMRI + 6 fMRI tasks
 - Variability **between tasks** ?
 - **Task paradigm** impact pipeline design
 - How well experiment finding on a task **generalize** to others ?
 - Use rest data for **fault injection** ?

Subject: 100206					
Download (Aspera client required)		Data Use Terms			
PROJECT: HCP-Young Adult 2025					
PACKAGES:					
Download	ID	Name	Files		
All					
+ queue	10715	Diffusion 3T Preprocessed Recommended	142		
+ queue	10716	Resting State fMRI 3T Preprocessed Recommended	89		
+ queue	10718	Structural Preprocessed Recommended for 3T and 7T	189		
+ queue	10719	Task fMRI 3T Recommended	375		
3 GB			9 GB		
10 GB					
1 GB					
9 GB					
10	/page	First	1 - 4 / 4 (16)		

balsa.wustl.edu/subject/46vPX

Neurovault

- > 16,000 collections
- Only analysis results (**unthresholded maps**)
- Interesting for **oracle testing and validation**
- Quality of data is **not guaranteed**



identifiers.org/neurovault.collection:16

NARPS

- fMRI data from **108 subjects** (anatomical + task-fMRI)
- Results (**unthresholded & thresholded maps**) were shared
- Pipelines were described, but **not shared**
 - **NARPS Open pipeline project**
 - Create a codebase **reproducing the 70 pipelines**
 - **Share it** as an open resource for the community

NARPS Open Pipeline

- Pipelines built with **NiPype**
- 7 pipeline fully implemented, ~45 in progress
 - e.g. `.../blob/main/narps_open/pipelines/team_Q6O0.py`
- **Evaluation** of results reproducibility
 - **Correlations** between shared maps and pipeline results

team_id	status	main software
O21U	●	FSL
3TR7	●	SPM
J7F9	●	SPM
L7J7	●	SPM
Q6O0	●	SPM
R9K3	●	SPM
U26C	●	SPM
27SS	●	AFNI
6WV2	●	AFNI

[.../wiki/pipeline_status](#)

Exploring the variability induced by different configurations in the neuroimaging analytical space

RIPOST Reproducibility Workshop - 2025-10-29

Thank you for your attention