

Research interests: data science, empirical inference, and brain image analysis.

- Drawing insight from data: machine learning, statistical modeling and data integration.
- Statistics on brain images to understand cognition.
- Methods and software for computational science: developing scientific libraries and good practices.
- Initial training in physics; PhD on quantum atom-optic.

EDUCATION

- HDR** Habilitation (*for full professorship*) in computer science, Sorbonnes Universités.
2018 Estimating brain functional connectivity and its variations from fMRI
- PhD** Université Paris Sud Orsay, FRANCE, direction: prof. Alain Aspect
2005–2007 Thesis: *Atomic sources for long-time-of-flight interferometric inertial sensors*
- ENS** Undergraduate studies at École Normale Supérieure Paris (ENS)
2001–2004 Masters in quantum physics (DEA de physique quantique, ENS)

POSITIONS

After PhD

- | | |
|------------------|--|
| Computer science | 2011-present INRIA, chargé de recherche de première classe <ul style="list-style-type: none">◇ Developed statistical learning and data analysis for cognition and psychiatry in brain imaging.◇ Developed and directed key software for machine learning and neuroimaging.◇ Research and outreach to democratize machine learning in data science applications. |
| | 2010-2011 INSERM, Unité de NeuroImagerie Cognitive (Unicog), Post-doc <ul style="list-style-type: none">◇ Clinical research on prognosis of the cognitive impact of strokes from fMRI. |
| | 2008-2010 INRIA, Post-doc <ul style="list-style-type: none">◇ Developed unsupervised methods to extract brain function from resting subjects.◇ Helped creating the PARIETAL INRIA team in NeuroSpin (brain imaging center). |
| Software | Summer 2008 UC Berkeley, programmer: nipy software for brain image analysis. |
| | Summer 2008 Enthout Inc, Austin Texas, software consultant <ul style="list-style-type: none">◇ Implemented data processing and visualization for scientific applications in the industry. |
| Physics | Fall 2007-Mid 2008 LENS (European Laboratory for Non-linear Spectroscopy), Italy <ul style="list-style-type: none">◇ Post-doc: atomic physics experiments (Bose-Einstein Condensation). |

IMPACT

Publications

h-index: 35 on Google scholar [↗](#)

Journal papers More than 50 peer-reviewed articles in journals such as NeuroImage, JMLR, IEEE TMI.

Conference papers More than 50 peer-reviewed conferences papers, in venues such as NIPS, ICML, MICCAI.

Software

- scikit-learn** scikit-learn project manager and developer: the reference machine learning Python package;
2009–Present more than 500 000 users.
- joblib** Author of joblib, light-weight and high-performance pipelining of scientific Python code.
2009–Present
- Nilearn** Contributing to and directing Nilearn: machine learning for brain imaging.
2012–Present
- Mayavi** One of the two developers of Mayavi, the lead 3D scientific data visualization Python tool.
2007–2014

DUTIES

Academic

Supervised
& co-supervised

- 10 PhDs defended, ■ 3 ongoing PhDs; ■ 9 post-docs,
- 3 junior software developers and 9 senior ones.

Teaching

- *Advanced machine learning*, 2015, Data-science master (Centrale Paris).
- *Brain functional connectivity*, since 2014, **Bio Medical Engineering** master (Telecom Paris).
- *Machine learning in Python*, since 2014, ENSAE.
- 40 hours a year of teaching scientific computing and data analysis in conferences and workshops.

Community service

- Head of the **scikit-learn consortium at Inria**.
- Associate director for the Paris-Saclay **Center for Data Science** since 2016
- Member of the Scientific Advisory board for Computer Science research at Total Inc.
- 2015–2017: **director of the joint lab Inria-tinyclues**: machine-learning for market analysis.
- Commission d'Évaluation Scientifique ANR 2016 (main French funding agency panel).
- 2013: **nominated to the Python Software Foundation**, that supervises the Python language.
- **Program chair** of IEEE Pattern Recognition in NeuroImaging 2013 (200 attendees).
- **General chair** of Euroscipy 2010 and 2011, program chair of Scipy 2008 (300 attendees).
- Member of **8 PhD committees** (3 outside France, 2 medical).
- INRIA Saclay: “Comité de suivi doctoral” and “Commission de développement technologique”.
- **Editor** (NeuroImage 2014–2017, Frontiers), Guest editor **J. Computational Science**, **reviewer** for international funding agencies.
- Editor of the **Scipy lecture notes**, an online book on scientific computing in Python.

Reviewer

in journals ranging from computer science to basic neuroscience (15 reviews a year), and major conferences (40 reviews a year): *Nature Methods*, *Nature Neuroscience*, *Nature Human Behavior*, *Human Brain Mapping*, *NeuroImage*, *Trends in cognitive science*, *NeuroInformatics*, *J. Physiology Paris*, *J. Machine Learning Research*, *J. Stat Soft.*, *Annal App. Stat.*, *Medical Image Analysis*, *Transactions in Medical Imaging*, *ICASSP*, *MICCAI*, *ICML*, *NIPS*, *ICLR*...

Major grants

	Funding	Amount	Period	Role
MissingBigData: missing data in the big-data era	DataIA	200 k€	2018-21	Co-PI
DirtyData: data integration and cleaning for statistical analysis	ANR	500 k€	2017-21	PI
INRIA-tinyclues lab: machine learning for market analysis	LabCom	300 k€	2014-16	PI
Wendelinia: big data for security in the Internet of Things	FUI	200 k€	2014-17	Co-PI
Niconnect: tools for clinical research with brain functional connectivity mapping	Investissement d'avenir	700 k€	2012-17	PI

RESEARCH

Data science and computing

All my research is anchored in mathematical modeling and algorithms, to draw conclusions from data.

Democratizing machine learning

- My new research project is about statistical learning on data without curation. It entails **data integration**, **knowledge representation**, **representation learning on databases**, learning with **missing data**.
- For 10 years, I have been developing machine-learning models used in production by non experts.

Computational and software aspects

To bring algorithm research to applications, I invest in building a general-purpose **tool stack for computational science using the Python language**, with many libraries and improvements to the ecosystem. I combine excellent software and numerical engineering with usability, documentation, and visualization considerations.

Brain image analysis

I have 10 years of experience using machine learning and multivariate analysis on brain images to understand brain function: relating brain measures to behavior, cognition and pathology. Beyond mathematical challenges, developing new inference methodology brings **epistemological questions**.

Mapping cognition Assigning specific functions to brain structures and decomposition behavior into cognitive processes can be seen as a **high-dimensional statistical-learning** problem. I have been interested in well-controlled machine-learning on brain images, as well as **accumulation of knowledge** across many studies.

Resting-state Neural activity at rest reveals brain structure that can form **biomarkers** of cognition or pathologies. Extracting them requires statistical modeling that combines **unsupervised and supervised learning** tailored to models of brain activity.

Physics

Research till 2008 **Atom optic** probes quantum behaviors of atoms. I have work on **interferometric measurements of inertial forces**, benefiting from atoms strong gravitational coupling. I have contributed to production of **quantum gases** with laser-cooled atomic sources. I have started the first atom interferometry experiment in **microgravity** using a freely-falling plane. For metrology, I have introduced **Bayesian statistical modeling** for tests of general relativity on a noisy platform.

SELECTED PUBLICATIONS

Statistical learning [1] Stochastic subsampling for factorizing huge matrices, A Mensch, J Mairal, B Thirion, G Varoquaux IEEE Transactions on Signal Processing, 66(1), 113-128, 2018, **6 citations**.

[2] Similarity encoding for learning with dirty categorical variables, P. Cerda, G. Varoquaux, and B. Kégl, B. (2018), Machine Learning, 1-18, 2018.

NeuroImaging methodology [3] NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. K. Gorgolewski, G. Varoquaux, G. Rivera, Y Schwartz, ... Frontiers in neuroinformatics, 9, 2015, **159 citations**.

[4] Machine learning for neuroimaging with scikit-learn, A Abraham, F Pedregosa, M Eickenberg, P Gervais, A Mueller, J Kossaifi, A Gramfort, B Thirion, G Varoquaux, Frontiers in neuroinformatics, 8, 14, 2014, **173 citations**.

[5] Brain covariance selection: better individual functional connectivity models using population prior. G Varoquaux, A Gramfort, JB Poline, B Thirion. NIPS, 2334-2342, 2010, **189 citations**.

Medical Imaging [6] Total variation regularization for fMRI-based prediction of behavior. V. Michel, A. Gramfort, G. Varoquaux, E. Eger, and B. Thirion. IEEE Transactions on Medical Imaging, 30, 1328-1340, 2011, **129 citations**.

[7] Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion, Information Processing in Medical Imaging-IPMI 562-573, 2011 **104 citations**.

Imaging Neurosciences [8] Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. A Abraham, MP Milham, A Di Martino, RC Craddock, D Samaras, B Thirion, G. Varoquaux, NeuroImage, 147, 736-745, 2017, **97 citations**.

Scientific computing software [9] Scikit-learn: Machine learning in Python. F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, ... The Journal of Machine Learning Research 12, 2825-2830, 2011, **12 794 citations**.

[10] The NumPy array: a structure for efficient numerical computation, S Van Der Walt, S Colbert, and G Varoquaux, Computing in Science & Engineering, 13(2), 22-30, 2011, **2931 citations**.