Gaël Varoquaux

Visiting scholar at McGill university / Research Director at Inria

Born on April 17th 1981, http://gael-varoquaux.info

My scientific interests explore how understanding emerges from data. I did a PhD on quantum-physics experiments before switching in 2008 to brain-image data analysis, in a computer science group. For this, I developed machine learning tools and progressively applied them to other problems motivated by industry and public-health partners, in particular machine learning on dirty data.

Since 2008, I have been working on applied statistics and machine learning. My driving question is: **how can machine learning answer "soft"-science questions?** I have first worked on informing psychology and psychiatry with brain imaging, notably using graphical models to quantify the activity of subjects at rest, or sparse methods for compressive-sensing like brain mapping with few measurements. I then focused on learning from aggregated data, to increase statistical power and build broader models from more diverse data. Beyond brain imaging, I am currently working on **facilitating machine learning on erroneous data**. My contribution include new learning formulations [3, 8], faster algorithms [1], and better signal processing [2, 5].

I have invested a lot in improving scientific computing with the Python language, to transfer algorithms outside of computer science. I have created or significantly contributed to very widely used software (Mayavi, 3D plotting; scikit-learnmachine learning; numpy , array computing; joblib, parallel computing).

Education

Habilitation	2018	University Paris Saclay, May 23rd 2018 in computer science
PhD thesis	2005-2008	Université Orsay, "Institut d'Optique" (optics department). *Atomic interferometry*; awarded on Jan 18th, 2008. Adviser: Alain Aspect
Master degree	2003-2004	Ecole Normale Supérieure Paris: Master in Quantum Physics
	2001-2004	Ecole Normale Supérieure Paris: degree in fundamental physics

Positions

Computer science / Software / Physics

2019-2020 McGill university, visiting scholar
- ◇ Brain imaging & mental health at MNI, Machine learning at MILA

2011-present Inria, tenured researcher
- ◇ Developed statistical learning to understand brain pathologies and cognition.
- ◇ Developed and lead key neuroimaging (nilearn) and machine learning (scikit-learn) software.
- ◇ Research and outreach to democratize machine learning in data science applications.

2010-2011 INSERM, Post-doc (medical research institution)
- ◇ Clinical research on data analysis for prognosis on the impact of strokes from MRI.

2008-2010 Inria, post-doc
- ◇ Developed unsupervised methods to study the brain activity of resting subjects.
- ◇ Helped starting a computer-science team in NeuroSpin (brain imaging center)

Summer 2008 UC Berkeley, software programmer, in the Brain Imaging Center

Summer 2008 Enthougt Inc, Austin Texas, software consultant
- ◇ Data processing and visualization for scientific applications (mainly in the oil industry)

Fall 2007 – Spring 2008 LENS (European Laboratory for Non-linear Spectroscopy), Florence, Italy, Research assistant and then Post-doc, atomic and quantum physics

Research Contributions and Impact. I have an H-number of 43 and my publications have been cited more than 42 000 times (both obtained from Google Scholar). See below for more details.

Teaching. I teach machine learning at ENSAE (major statistics graduate school in France) and brain-image analysis at the Paris Bio-Medical Imaging master.

PhD supervision. I have supervised and co-supervised 10 students who have defended; I am currently supervising 3 PhD students.machine learning in uncurated tables, and machine learning with missing data.

Other supervision.

- I have supervised **8 post-docs**
- For machine learning and neuroimaging softwares, I have supervised **5 junior software developers**, and **7 senior software developers**.

Responsibilities

Director of the scikit-learn consortium Scikit-learn is one the **leading machine-learning library** with more than 1000 contributors, 25 000 citations, used very widely in the industry and in academia. I initiated the library in 2010, have been contributing to it, managing the community and industry relations.

Director of the joint lab INRIA-tinyclues (2014–2016). Tinyclues is uses machine learning for business analytics. The lab employed 2 engineers to develop new learning methods for market analysis.

Community service

Conference chair : **Senior Program Committee** member IJCAI, **General chair** of Euroscipy 2009 and 2010 (200 attendees), as well as **program chair** of IEEE Pattern Recognition in NeuroImaging 2013 (200 attendees, IEEE proceedings) and **steering committee chair** 2014 - 2015, IEEE PRNI.

Nominated member of the *Python Software Foundation* (since 2013), that supervises the community around the Python programming language.

Paris-Saclay Center for Data Science associate director

Academic publishing **Editor** at elife (very high impact journal for computational biology); previously editor at NeuroImage from 2014 to 2017 (main journal of the brain imaging community), Frontiers in Brain Imaging methods, Frontiers in NeuroInformatics. **Reviewer** for funding agencies (ANR, NWO) journals (15 reviews a year), and conferences (40 reviews a year).

Grant panel Commission d'Évaluation ANR 2016: main scientific funding agency in France

Industry advisory I am member of the Scientific Advisory Board for computing at Total (world-wide petrol company) and scientific consultant for St Gobain (world-wide construction-material company).

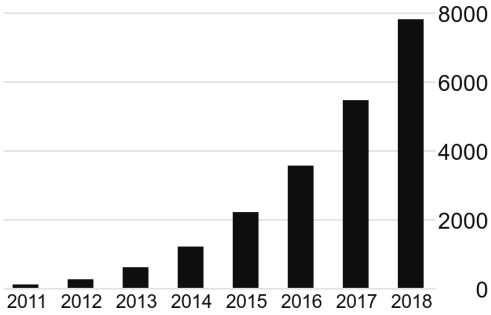
Juries 4 French PhDs, 1 MD-PhD, 3 International PhDs, 1 faculty-recruitment committee.

Institutional committees Since 2013: member of the graduate-student advisory committee and the technical development committee of Inria Saclay.

Scientific track Record

My h-index is 43 on Google scholar. As conference proceedings are important in my field, I use Google scholar.

I publish in machine learning and brain-image processing. In each of these fields, I aim for the best venues, often conferences in computer science.
My publications can be found on my Google scholar page.



Major grants.

Name and topic	Funding	Amount	Period	Role
Niconnect: tools for clinical research with brain functional connectivity mapping	Investissement d'avenir	700 k€	2012-17	**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
DirtyData: data integration and cleaning for statistical analysis	ANR	500 k€	2017-21	**PI**
MissingBigData: missing values in the era of big data	DataIA	200 k€	2018-21	Co-PI
LearnI: learning data integration	Chair IA	600 k€	2020-24	**PI**