



COLAS DROIN

Data scientist

📍 Geneva – Switzerland

✉️ French (Permit C)

⚡ 31 years old

✉️ colas.noe.droin@cern.ch

📞 +41 76 651 83 05

PROFILE

Data scientist with a background in computational physics, my main interests include data visualization, mathematical modelling, machine learning, web and software development.

I am aiming to work as a data scientist with a heavy emphasis on data visualization.

EDUCATION

PhD in computational biophysics

École Polytechnique Fédérale de Lausanne (EPFL)

2016 – 2020

MSc in computer science and physics of complex systems

École Normale Supérieure (ENS) de Lyon

2015 – 2016

MS in biomathematics and bioinformatics

Institut National des Sciences Appliquées (INSA) de Lyon

2013 – 2015

SOCIAL

linkedin.com/in/colas-droin/

github.com/ColasDroin

tinyurl.com/snmxhbhy (scholar)

colasdrioi.github.io (portfolio)

EXPERIENCE

SENIOR FELLOW – SCIENTIFIC PROGRAMMER

Jan. 2023 – Now | Accelerators and Beam Physics Group, CERN

I am engaged in the development and maintenance of Python packages for the conception, parametrization and execution of Large Hadron Collider (LHC) simulations. This involves the coding of testing suites, Docker containerization for HPC cluster submission, and CI/CD pipelines to ensure robust and regular software delivery.

I also developed a web-based dashboard¹ for the monitoring and visualization of collider observables, which is now used by the group to assess a collider viability before launching simulations.

DATA VISUALIZATION SPECIALIST

Apr. 2021 – Oct. 2022 | Neurodevelopmental Systems Bio Lab, EPFL

I developed a set of visualization and machine learning tools in the form of a Python-Dash web-application², to turn a raw MALDI imaging dataset (~300GB) into a multidimensional atlas of rodent lipid brain composition. This involves data wrangling, analysis with dimensionality reduction and clustering methods, and data visualizations.. The app is now actively used in the lab to explore the dataset and generate hypotheses.

RESEARCHER (PHD AND POST-DOC)

Sep. 2016 – Dec. 2020 | Computational Systems Biology Lab, EPFL

I worked on multiple computational biophysics projects having for theme *Single-cell analysis of noisy biological oscillators and their interactions*³, and was awarded the **EPFL 8% PhD Thesis Distinction**⁴ as a result. These projects are described in my portfolio⁵.

- I conceived and optimized a Hidden Markov Model to better understand the non-linear dynamics of the cell-cycle and the circadian clock.
- I studied a dataset of spatiotemporal liver mRNA expression using mixed-effect models representing the Fourier harmonics of the system.
- I optimized a RNA velocity inference method based on differential geometry to quantify cell-cycle evolution from experimental data.

I implemented all these analyses in Python, and ran them on the EPFL HPC platform (SCITAS) when needed. All the code is open-source.

MASTER THESES

Spring 2015 and 2016 | IXXI (Lyon) and Dassault Systems (Paris)

- ENS MSc THESIS (2016): Extraction and analysis of information cascades in a large network graph built from a Twitter dataset. (R, Python)
- INSA MS THESIS (2015): Study and development of a hybrid system (deterministic-stochastic) for the Chemical Master Equation. (C++, R)

FIRST-AUTHOR PUBLICATIONS

Space-time logic of liver gene expression at sublobular scale, C. Droin, J. El Kholtei, K. B. Halpern, F. Naef, S. Itzkovitz, *Nature Metabolism*. 2020.⁶

Low-dimensional Dynamics of Two Coupled Biological Oscillators, C. Droin, E. Paquet, F. Naef, *Nature Physics*. 2019.⁷

AWARDS

- First prize, WEkEO hackaton
2500€ cash prize (2022)⁸
- First prize, CERN webfest (2022)⁹
- Amongst top cited paper of₁₀
Nature Metabolism (in 2021)
- EPFL Outstanding PhD Thesis
Distinction, Top 8% (2020)⁴
- Organizers prize, LauzHack 5.0,
Lausanne (2020)
- Best poster award, DDay, CHUV
Lausanne (2019)
- Best poster award, EPFL Physics
days (2018)

COURSES

Relevant PhD courses taken:

- EE-559 Deep learning
- CS-433 Machine Learning
- COM-480 Data viz
- Better coding practices (Swiss Institute of Bioinformatics course)

LANGUAGES

: fluent : conversational

LEARNING

Things I'm learning or want to learn:



INTERESTS



Check the online version of this CV
at: colasdroin.github.io/cv/

DATA SCIENCE SKILLS

My background involves a lot of statistical learning from data. I have used a broad spectrum of machine learning methods, including deep learning, and worked with large datasets. I'm also very keen on data visualization and the implementation of interactive representations. Python tools/libraries I've used: Jupyter, Poetry, Pytest, Numpy, Scipy, Numba, Pandas, Matplotlib, Seaborn, Plotly, Dash, Scikit-learn, Pytorch, Statsmodels, ipywidgets, Jinja2.

SOFTWARE ENGINEERING SKILLS

I am an experienced Python developer (8y+), used to work with version control (git), unit testing (pytest), and CI/CD pipelines (gitlab). The need for submitting large-scale simulations on HPC clusters has led me to learn about SLURM and HTCondor, along with Docker for reproducibility, and REST APIs for fetching data. I have also been exposed to R (MS thesis), Matlab (as a TA), Julia (PhD) and C++ (MS thesis) in the past.

As an OS X user, and having developed mainly on Linux clusters, I'm familiar with Unix systems and bash commands.

WEB DEVELOPMENT SKILLS

I've learnt web-development through the creation of several web-apps^{1, 2, 11} with Python Dash. I've also developed my own personal website with React⁵ and this CV¹² is made with HTML/CSS. My knowledge of JS frameworks remains limited, but I'm actively learning (cf. left panel).



Dash (Plotly)



HTML



CSS



Javascript



TEACHING AND TUTORING

Autumns 2017–2019 | Dynamical systems in biology EPFL: ,

As the head TA of this 3rd-year bachelor course, I managed the exercises sessions. I rewrote the old Matlab problems into interactive Python Jupyter notebooks. I participated to the conception and grading of homeworks.

Spring 2017 | Probabilities and statistics II EPFL: ,

I was helping students with the exercises of this 2nd-year bachelor course.

Autumn 2012 – Spring 2014 | General tutoring, INSA Passerelle Program:

I provided assistance to 1st-year bachelor students with learning disabilities.

REFERENCES

Felix Naef

Full Professor

+41 21 693 16 21

felix.naef@epfl.ch

Gioele La Manno

Associate Professor

+41 21 693 84 46

gioele.lamanno@epfl.ch

Guido Sterbini

CERN staff physicist

- undisclosed -

guido.sterbini@cern.ch

LINKS

1: <http://tinyurl.com/t3u5bd9b>

2: lbae.epfl.ch

3: tinyurl.com/f34tfvwvp

4: tinyurl.com/y343w7eu

5: colasdroin.github.io

6: tinyurl.com/4rupr436

7: tinyurl.com/yhex8d72

8: tinyurl.com/2wvzcwjm

9: <http://tinyurl.com/5n63bnza>

10: tinyurl.com/2p8fbamv

11: czviz.epfl.ch

12: tinyurl.com/yvmcw5hf