Master Internship (Stage de fin d'étude) — Historical Wildfire Data for Conventional and AI-based Fire Spread Simulation, Prediction and Parameter Calibration (with potential PhD position)

Context

Historical wildfire observations are essential for evaluating fire spread simulators and calibrating their parameters. However, fire perimeters, ignition information, weather records, topography and vegetation data are scattered across different sources, with inconsistent formats, resolutions and coverage. Assembling these inputs remains a major obstacle to reproducible comparisons between simulators.

This project aims to develop a reproducible data pipeline for historical large wildfires in France and Europe. The resulting datasets will support comparisons between conventional simulators, such as FlamMap, and PyTorchFire (Xia and Cheng, 2025), a differentiable, GPU-accelerated wildfire simulator. They will also enable parameter calibration against observed fire behaviour. The EPHE team led by Christelle Hély will co-supervise the project, contributing existing fuel and vegetation measurements across seven ecosystem types in several French regions. These measurements and the team's curated historical fire records will support the evaluation and calibration of FlamMap and PyTorchFire

Research Objectives and Approach

The goal is to develop a Python pipeline (similar to FireDataForge (Xia et al., 2026) for US wildfires) that retrieves and harmonises data for selected historical wildfires, producing documented event datasets suitable for simulation and calibration. The work will include:

- Data retrieval and event selection: Identify suitable historical fires and combine EPHE field measurements of fuel loads and vegetation structure with EFFIS, Copernicus EMS mapping, French fire records, satellite observations, ERA5-Land weather, and terrain and vegetation products.
- ML-ready pipeline and quality control: Automate retrieval, spatial and temporal alignment, feature extraction and dataset export. Document provenance, native resolutions, missing values and uncertainty in ignition locations and observation times.
- Simulator interfaces: Prepare inputs for PyTorchFire and FlamMap, documenting mappings from vegetation, land cover and biomass data to simulator-specific fuel properties. Align ignition and weather scenarios with each simulator's capabilities.
- Benchmarking and parameter calibration: Compare simulated and observed burned areas, perimeters and fire progression where sufficiently resolved observations are available. Explore gradient-based calibration with PyTorchFire, regional parameter regression and an optimisation baseline for FlamMap, evaluating accuracy, computational cost and generalisation on independent test fires.

Expected outcomes

- A reusable Python pipeline and a curated collection of machine-learning-ready historical wildfire datasets, with documentation and reproducible examples.
- A comparative benchmark of FlamMap and PyTorchFire, assessing the benefits of biomass information and exploring data-driven parameter calibration and data assimilation using wildfire observations.
- A technical report, with potential contributions to a scientific publication and an openly released data pipeline, subject to source licences.

Potential continuation as a PhD to map wildfire burn probabilities across France (with similar approach to Palaiologou, P. et al., 2026) using calibrated simulation ensembles with randomly sampled ignitions, and to learn region-specific fire parameters through regression, comparing and interpreting their variation in relation to biomass, vegetation and climate. Continuation on PhD will depend on internship performance and mutual agreement.

Profile required

- Master 2 or final-year engineering student in Computer Science, Computational Geoscience or a related field.
- Strong Python (Pytorch) programming skills and experience with scientific data processing.
- Experience with geospatial data, GIS or remote sensing would be a strong asset.
- Experience with Flammap and PyTorchfire would be appreciated.
- Interest in wildfire modelling and environmental applications, self-motivation, and very good written and spoken English.

Supervision and environment

The internship will take place at CEREA, ENPC / Institut Polytechnique de Paris, under the joint supervision of Dr Sibo Cheng and Christelle Hély (École Pratique des Hautes Études, EPHE). The project is funded by QWERTY within the PEPR FORESTT programme.

Practical information

- Duration: 5–6 months (in spring 2027).
- Location: ENPC, Champs-sur-Marne, with possible remote work up to 1-2 days per week.
- Compensation: standard ENPC internship allowance and transport support.

Application

Send a single PDF containing a CV, a short motivation letter, Master's transcripts and contact details for two academic referees to sibo.cheng@enpc.fr and christelle.hely@ephe.sorbonne.fr . Please indicate your interest in continuing towards a PhD. Applications may be written in English or French.
Email subject: Application – Wildfire Data Internship – Your name

References (recommended to read before application)

Xia, Z., Chen, L., Liu, Y. and Huang, H. (2026). FireDataForge: A Unified Framework for Multi-Source Wildfire Data Retrieval and Integration. In 2026 IEEE International Conference on Information Reuse and Integration for Data Science (IRI), pp. 528–531. https://doi.org/10.1109/IRI69576.2026.00103.

Xia, Z. and Cheng, S. (2025). PyTorchFire: A GPU-accelerated wildfire simulator with Differentiable Cellular Automata. Environmental Modelling & Software, 188, 106401. https://doi.org/10.1016/j.envsoft.2025.106401.

Palaiologou, P. et al. (2026). A dataset to support wildland fire and fuel management in Greece created with stochastic wildfire simulations. Data in Brief, 64, 112304. https://doi.org/10.1016/j.dib.2025.112304.