Willian Vieira PhD

Geospatial Data Engineer & Scientist with a PhD in Ecology, I integrate AI/ML models with cloud-native data engineering to build scalable, automated pipelines for environmental impact.

## 💼 DATA SCIENCE & ENGINEER EXPERIENCE

2025 | 2024 (1 yr 9 m)

Data Analyst
Habitat, Montreal, Canada

Built and maintained reproducible pipelines using satellite imagery and open-source spatial data for automated monitoring | Designed a geospatial lakehouse architecture for streaming unstructured datasets from Azure cloud storage, optimizing data retrieval for analytics | Architected end-to-end R/Python ETL pipelines to deploy geospatial ML models into production environments | Implemented metadata-driven workflows and data validation checks, significantly improving pipeline reliability and governance | Led the transition to a Unix-based production environment, creating automated setup scripts and modernizing workflows with Docker, CI/CD, and DevOps best practices

2024 | 2017

PhD Research
Integrative Ecology Lab, Sherbrooke, Canada

Built end-to-end geospatial data pipelines to ingest, clean, and harmonize continental-scale forest-climate datasets for Bayesian hierarchical modelling | Developed reproducible HPC (High-Performance Computing) workflows (⊙), leveraging parallel processing and cluster management to execute thousands of model simulations at scale | Developed open-source software libraries (⊙, ⊙) for geospatial demographic modelling, implementing version control and automated testing | Published a technical methods book documenting the full computational pipeline to ensure long-term project reproducibility, along with one peer-reviewed publication and two preprints (📖, 📖)

2022 | 2020 (part-time)

Biostatistician
Environment and Climate Change Canada - Quebec, Canada

Developed a cost-aware probability sampling protocol to optimize the spatial representativeness of boreal bird surveys in Quebec | Led R&D on spatial bias correction methodologies, designing a method later adopted by other provinces | Engineered a fully automated and reproducible geospatial pipeline featuring automated documentation and version-controlled workflows to ensure long-term system maintainability | Produced a fully open source, reproducible analytical workflow with automated documentation

🎓 EDUCATION

2024 | 2017

PhD, Ecology
Université de Sherbrooke - Sherbrooke, Canada

📄 *How climate, competition, and forest management shape the limits of tree species distributions: from individuals to metapopulations*

2016 | 2015

Masters 2, Agroecology and Resource Management
Bordeaux Sciences Agro, Bordeaux, France

📄 *Modelling the dispersion of weed species in agricultural landscapes*

2015 | 2010

BSc in Agronomy
Universidade Federal de Santa Catarina, Florianópolis, Brazil

---

CONTACT

📞 (263) 381-3636
✉ w.vieiraw@gmail.com
in Linkedin/WillVieira
⊙ GitHub/WillVieira
✇ Publications

SKILLS

Data Engineering & Cloud:
ETL/ELT · Azure · AWS · Docker · Apache Arrow · Airflow & DBT · COG · GeoParquet · HPC

Programming & Database:
Python · R · Julia · DuckDB · PostGIS · Bash

DevOps & Workflow:
Git · CI/CD · Unix · Unit testing · Make · Quarto · Jupyter

Statistics & Modelling:
Bayesian inference · Machine Learning · Hierarchical models · Experimental design

OUTREACH

· Developed and maintained the QCBS R Workshop Series, reaching nearly one thousand graduate students · Taught +250 students across biology, engineering, and graduate programs (programming, statistics, ecology) · 5 published papers & preprints with ~100 citations · Translated client needs into technical analyses and delivered clear, actionable results

LANGUAGES

Portuguese · Native
French · Full Professional
English · Full Professional

CV source code hosted on ⊙
Last updated: 2026-02-15
⤓ *PDF download available*