

Willian Vieira PhD

Applied ML scientist with a PhD in quantitative ecology. I build statistical models, production-minded ML pipelines, and reproducible systems that turn noisy historical data into validated algorithms and decisions people can trust.

DATA SCIENCE & ENGINEER EXPERIENCE

2025
|
2024
(1 yr 9 m)

● Data Analyst & developer






Habitat, Montreal, Canada

Built production-minded R/Python analytical pipelines that turned messy real-world project data into reusable, validated model inputs and decision-ready outputs | Prototyped and iterated data products with historical/project data, using validation checks and stakeholder feedback to improve reliability and usefulness | Designed metadata-driven ETL and provenance workflows so datasets, definitions, analyses, and handoffs remained traceable, reproducible, and inspectable | Led the transition to a Unix-based production environment using Docker, CI/CD, automated testing, documentation, and collaborative development practices | Designed cloud/Azure-oriented data infrastructure that made heterogeneous analytical datasets easier to retrieve, validate, and reuse

2024
|
2017

● PhD Research

Integrative Ecology Lab, Sherbrooke, Canada

Developed Bayesian hierarchical models, simulation workflows, and historical-data evaluation routines to forecast forest dynamics from noisy, biased, incomplete real-world data, with explicit uncertainty propagation and model comparison | Ran thousands of simulations on HPC infrastructure () to test algorithmic assumptions, compare model behavior, and scale analyses beyond local compute | Built reproducible pipelines for continental-scale geospatial and environmental data, harmonizing climate, satellite, and field-inventory datasets into modeling-ready inputs | Developed open-source software libraries (, ) for demographic modelling with version control, automated testing, documentation, and reusable workflows | Published a **technical methods book** documenting the full computational pipeline, 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 improve the spatial representativeness of boreal bird surveys in Quebec | Led R&D on spatial bias-correction methods for large-scale ecological monitoring data, designing an approach later adopted by other provinces | Engineered a fully automated and reproducible **data pipeline** with version-controlled workflows, automated documentation, and stakeholder-ready analytical outputs

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
 [LinkedIn/WillVieira](#)
 [GitHub/WillVieira](#)
 [Publications](#)

SKILLS

Machine Learning & Statistics:

Bayesian inference · Hierarchical models · Optimization · Model evaluation · Simulation · Feature engineering

Programming & Database:

Python · R · SQL · Bash · DuckDB · Stan · Julia

Production ML & Data Systems:

Production ML pipelines · ETL/ELT · APIs · Docker · Azure · AWS · Apache Arrow · HPC

DevOps & Reproducibility:

Git · CI/CD · Unix · Unit testing · Reproducible pipelines · Automated validation · Provenance

OUTREACH

· Developed and maintained the **QCBS R Workshop Series**, reaching nearly one thousand graduate students · Taught programming, statistics, SQL, and reproducible workflows to 250+ students across biology, engineering, and graduate programs · Translated stakeholder needs into technical analyses, documented pipelines, and clear analytical outputs

LANGUAGES

Portuguese · Native
French · Full Professional
English · Full Professional