habitat
LA NATURE À L'ŒUVRE

Who is Willian Vieira?

and what has he been doing?

Willian Vieira, Analyst

WV-Habitat/me

@WillVieira90

Outline

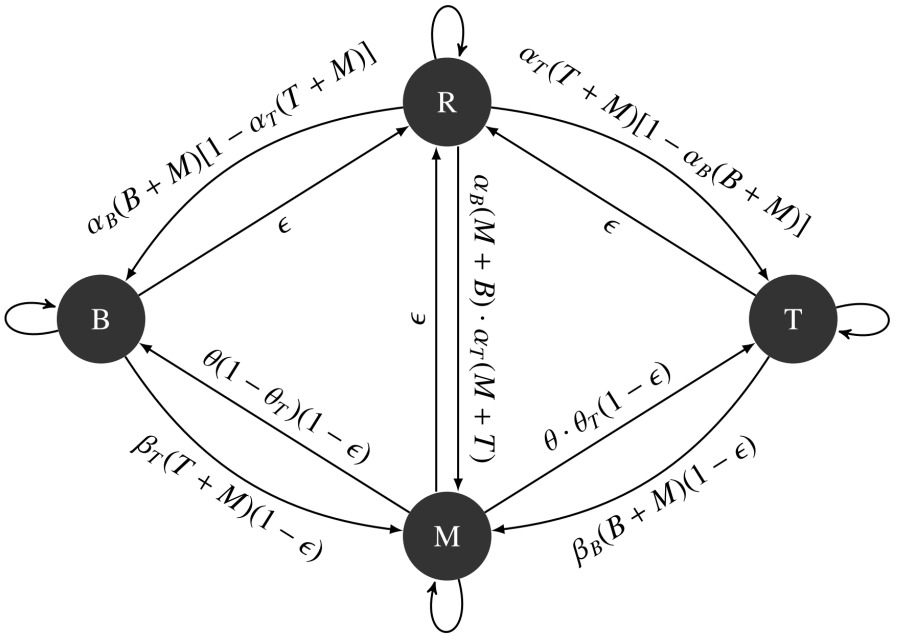
- My thesis project
- The ECCC project
- Side projects
- Future projects

My PhD thesis

How can we better predict tree species distribution?

- Species Distribution Models (SDM) are not appropriated for trees
- But we can use *theory* to create **mechanistic models**
- What are the drivers of forest dynamics shaping tree distribution?
- Mathematical and statistical models to assess the effect of climate, competition, and forest management

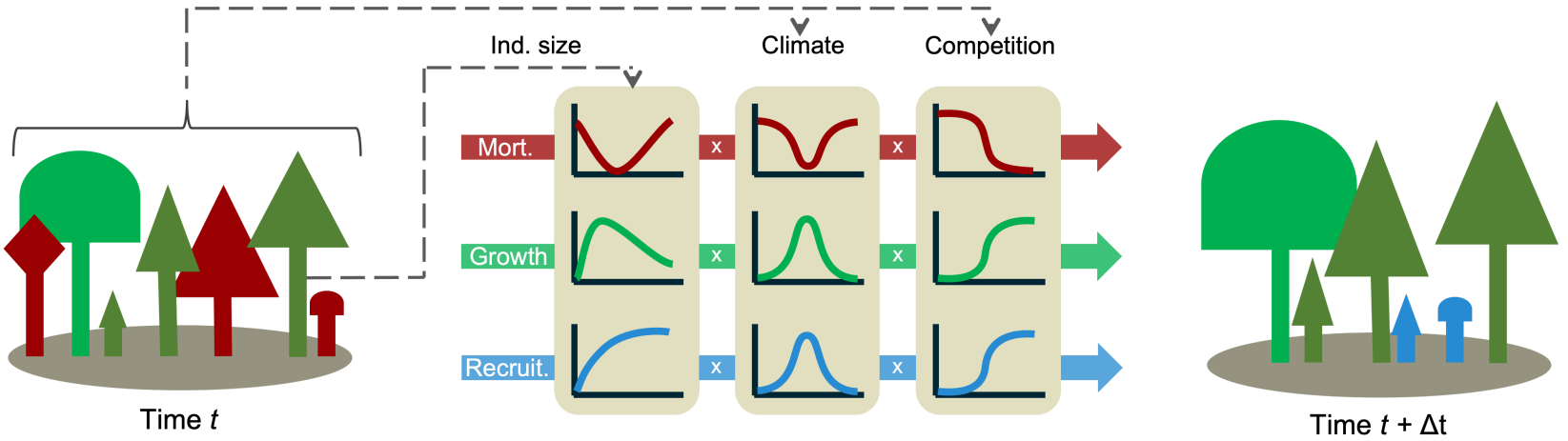
Chapter 1: State Transition Model



- Derived from the metapopulation theory
- **B**oreal, **T**emperate, **M**ixed, and **R**egeneration ~ MAT + MAP
- Extension to include forest management practices
- Simple model -> analytically tractable -> transient dynamics
- Matrix algebra, Jacobian approximation
- STM R package; Shiny App

Model from Vissault et al. 2020

Chapters 2 & 3: Integral Projection Model



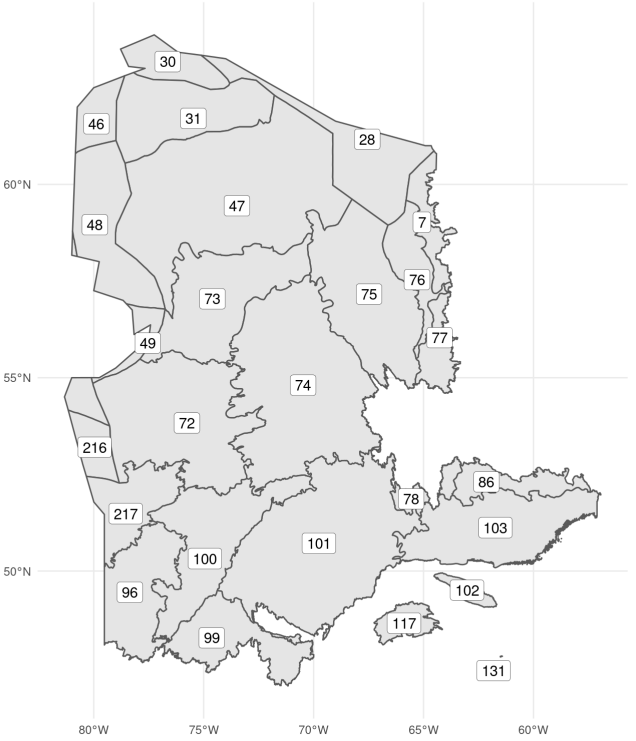
- Demographic models for 31 tree species from eastern North America
- λ ~ [Growh + mortality + recruit] ~ climate + competition
- Non-linear hierarchical Bayesian models
- Stan: statistical programming language derived from C++
- High-performance computing (HPC)
- C.2: Random forest for sensitivity analysis
- C.3: Scale integration - from individual performance to the metapopulation

The ECCC project

Sampling design for monitoring boreal birds in Quebec

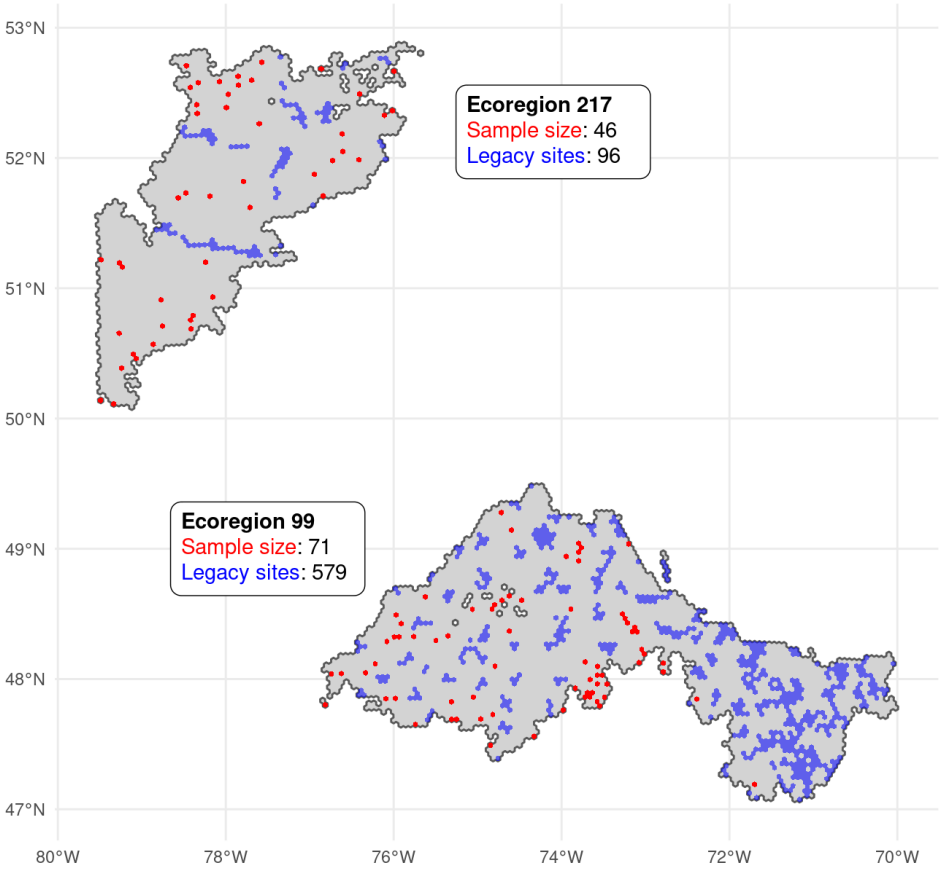
BMS project

- National effort to develop a protocol to sample boreal birds
- Adapt and implement national design for the Quebec region
- Spatially stratified sampling approach weighted by:
  - Habitat
  - Cost
  - Legacy sites
- Report



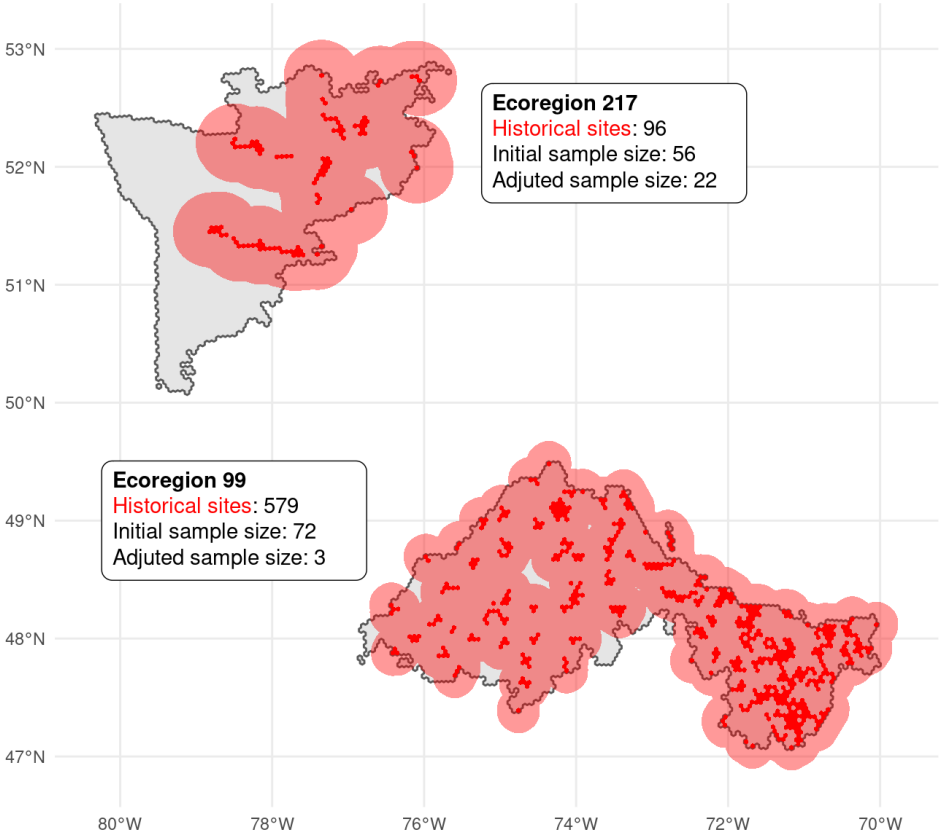
BMS project

R&D: A new approach to account for legacy sites



BMS project

R&D: A new approach to account for legacy sites



Side projects

From learning useless programming languages to automating unnecessary tasks

Side projects

Reporting and templates

- Manuscript template (using markdown, LaTeX, Pandoc, Lua and yaml)
- Presentation template (this presentation, some CSS and JavaScript)
- Books (Lab notebook, ECCC report)

Task automation

- Make (reproducibility of manuscripts)
- Crontab (Kijiji scraper R package)
- GitHub Actions to automatically:
  - Test and build R packages
  - Reproduce analysis
  - Deploy reports and websites

Future projects

Or what I would like to explore

Future projects

Modeling

- Statistical models (spatiotemporal autocorrelation, time series)
- Machine learning beyond random forest

Development tools

- Modularization (Packages, unity test, deployment, versioning, code review)
- Containerization (Docker)
- Data Engineering
- Cloud computing

Cloud-based geospatial technologies

Cloud-based geospatial technologies

Cloud Optimized GeoTIFF (COG)

It is based on two complementary frameworks:

1. Optimal GeoTIFF storage (store and organize each pixel information)
    - Tiling
    - Overviews
2. HTTP GET range requests
    - extract only a portion of the GeoTIFF file
    - Not mandatory but already built into cloud services (Google, Azure)
    - Ranges are determined by external metadata

Cloud-based geospatial technologies

Spatial-Temporal Asset Catalog (STAC)

- A common language to describe geospatial information
- metadata standard (JSON)
- Structured metadata repository describing
  - **What** it is
  - **Where** data is located
  - **How** it can be used
- Hierarchically structured into items, collections, and catalogs

Example with the Landsat images:

- RED band of an image is an asset
- All color bands of an image is an item
- All images together are a collection

Cloud-based geospatial technologies

Why should we move towards STAC?

- Once linked to STAC, we have access to any [new] open data effortless
- Catalog is centralized, providing easy searchability of new data sets
- Inclusion of new data is easy
- Community-based metadata extensions for specific problems

Constrants

- It comes with high implementation costs (especially if we become data providers)
- Less useful if we are interested in only a few spatial datasets