COMP 7980 — Capstone Project

Spatiotemporal Prediction of Hurricanes and Forest Attributes

Co-instructors: Li An and Zhaofei Fan | Semester: Spring 2027

Course Description and Objectives

This computational research capstone integrates historical U.S. hurricane records and national forest inventory/health observations to develop a reproducible modeling pipeline and a prototype query tool. For a user-selected location (from a broad region to an approximately one-acre site) and date/time (year, month, day, or hour), the intended outputs are (1) an estimated probability of hurricane occurrence or exposure within a defined time window and intensity threshold, and (2) estimated forest attributes, such as tree species composition, growing-stock volume, biomass, mortality/damage indicators, and invasive plant occurrence where supported by the available data. The student will combine geospatial data engineering, statistical modeling, machine learning, uncertainty estimation, and scientific visualization.

The project explicitly distinguishes historical climatological risk from short-term operational hurricane forecasts. Fine temporal or one-acre outputs are exploratory model estimates—not direct observations—and must be accompanied by clear spatial/temporal support, calibration, and uncertainty. Forest variables observed intermittently in FIA/FHM must not be presented as measured hourly conditions.

Expected Learning Outcomes

- Understand variable definitions, sampling design, coverage, spatial precision, and limitations of the FIA/FHM and hurricane datasets.
- Formulate operational definitions for hurricane probability (location, intensity threshold, exposure radius, forecast/historical window) and forest-attribute targets.
- Build an automated pipeline to download, standardize, audit, and integrate multiple large spatiotemporal datasets.
- Develop and validate probabilistic hurricane-risk models and spatial/temporal forest-attribute prediction models.
- Quantify model accuracy, calibration, geographic transferability, scale effects, missing-data sensitivity, and uncertainty.
- Deliver reproducible code, maps, a query-based prototype, and a research report explaining strengths and limitations.

Research Datasets

Source	Role and main variables	Data/access
USFS FIA Phase 2	Permanent plots, site condition/tree tables; species, volume, biomass, growth, mortality, disturbance, diameter, etc.	FIA DataMart: https://research.fs.usda.gov/products/dataandtools/fia-datamart
USFS FHM / Phase 3	Health indicators on subsets of plots; e.g., crowns, soils, down woody	FIA / forest health guides: https://www.fs.usda.gov/science-technology/forest-health-protection/monitoring

	material, selected invasive/understory measurements (availability varies).	
NOAA HURDAT2	Best tracks: storm ID, time, center coordinates, winds, pressure, available wind radii.	https://www.nhc.noaa.gov/data/
NOAA IBTrACS	Cross-agency tropical cyclone tracks/intensity, storm-time geospatial tables.	https://www.ncei.noaa.gov/products/international-best-track-archive
NOAA Storm Events	U.S. reported events and losses: dates, places, property/crop damages, fatalities and injuries.	https://www.ncei.noaa.gov/access/storm-events-database/

Supplemental predictor layers (e.g., climate, topography, soils, land cover, remote sensing, and gridded wind/rainfall) may be evaluated if available and approved. Hurricane records alone do not provide exact local wind footprints; Storm Events damages are not measurements of forest attributes.

Major Course Components

- In-class meetings: Six approximately two-hour project meetings, held approximately every three weeks, for research discussions, code reviews, and presentations. Time/location to be arranged by the co-instructors. Attendance and participation are mandatory.
- Progress reports: Ten approximately 30-minute meetings (typically on weeks without in-class meetings), online or in person as arranged. Students submit a concise written progress log with obstacles, decisions, and next steps.
- Reading and article presentations: Early readings on FIA/FHM design, hurricane datasets, probabilistic forecasting, spatial modeling, and uncertainty; two student-led presentations.
- Hands-on computing: Python/R, Bash, GitHub, GIS and, where approved, Auburn University HPC resources; reproducible versioned workflows and peer audit.

Prerequisites

Designed for graduate students with programming experience in Python or R and familiarity with basic statistics. Prior GIS, Bash, and machine-learning experience are recommended; relevant training will be incorporated at the start of the course.

Grading Criteria

Activity	Counts	Points/count	Total
Class attendance and discussion	6	25	150 (15%)
Progress reports	10	10	100 (10%)
Article presentations	2	50	100 (10%)
Coding tasks	5	90	450 (45%)

Final project report / presentation / repository	1	200	200 (20%)
Total			1,000 (100%)

Letter grades: A ≥90%; B ≥80% and <90%; C ≥70% and <80%; D ≥60% and <70%; F <60%.

Late Policy (Coding Tasks)

- Students have two no-questions-asked late days to use during the semester.
- After those days are used, each late coding task incurs a 20% deduction per day. Because code auditing is a joint responsibility, the penalty can apply to both coding and auditing work when either party causes delay.
- Tasks submitted more than five days late will not be accepted because subsequent tasks depend on them. Exceptions follow university policy for approved circumstances.

Assessment Details

1. Class Attendance — 150 points

Six meetings × 25 points; includes attendance, preparation, discussion, and demonstration of project progress.

2. Progress Reports — 100 points

Ten reports × 10 points. Before each meeting, provide completed work, a code/data change log, results, problems encountered, and planned actions; document feedback afterward.

3. Article Presentations — 100 points

Two presentations × 50 points. Evaluation: clear synthesis of research questions and findings (40%), critical assessment of methods and limitations (30%), and presentation/discussion quality (30%).

4. Coding Tasks — 450 points

Five tasks × 90 points. All scripts and documentation are managed via GitHub. Each task must be audited by a peer within 48 hours after submission; the audit includes running representative code, checking outputs, and identifying assumptions or errors.

Task	Focus/deliverable	Key grading requirement
1	Data acquisition and reproducible setup; Bash/Python/R download scripts, data dictionary and manifest	Automated downloads, stable source references, version control and provenance; protected/restricted FIA information handled lawfully.
2	Spatiotemporal harmonization and integration; common schemas, spatial joins, linked analysis table	Correct storm IDs/times/units, coordinate systems, duplicate handling, FIA table joins and documented linkage constraints.
3	Quality control and exploratory analysis; reproducible QC scripts and maps	Missingness/outlier diagnostics, storm coverage and reporting bias, FIA remeasurement gaps, geographic representativeness.

4	Feature engineering and preliminary models; hurricane risk and forest attribute features	Clearly specified probability event/window, spatial exposure proxies, lagged disturbance and environmental predictors, baseline model performance.
5	Modeling, validation and query prototype; scripts, figures and demo	Calibrated probabilities, forest-model performance, spatial/temporal holdouts, uncertainty, scale sensitivity and functional example queries.

5. Final Project — 200 points

Final written research report and methodology (60 points); oral presentation and demonstration (60 points); final GitHub repository with documentation, environment requirements, reproducibility and peer-audit resolution (80 points).

Optional bonus (maximum 100 points): up to 70 for exceptional peer review and up to 30 for clear, substantive code comments. Bonus does not raise the course total above 1,000 points.

Core Modeling Questions and Expected Prototype

1. Hurricane module: Given a location, reference date/time, temporal window, and storm/intensity definition, estimate the historical or model-based probability of a hurricane passing within a specified distance or exposing that location to a specified hazard. Compare suitable baselines, assess calibration, and distinguish climatological probability from an actual near-term forecast.
2. Forest module: Given a location, target year/time, spatial support, and selected variable, predict forest-cover status and supported attributes (e.g., species group/composition, live-tree volume, biomass, mortality, damage, and available invasive-species indicators), with uncertainty and explicit handling of nonforest locations and unobserved dates.
3. Integration module: Generate hurricane-history exposure variables for forest observations, explore how hurricanes may influence subsequent forest attributes, and compare models with and without hurricane-related features. Avoid asserting causality from predictive associations alone.
4. Query interface: Accept year/month/day/hour and region/county/location/approximately one-acre polygon, return modeled values and confidence/credible or prediction intervals (as appropriate), spatial resolution, data support, and warnings where inputs exceed model validity.

Feasibility boundaries: FIA plots are sparsely sampled and remeasured every several years, and public plot coordinates may be intentionally altered. One-acre predictions require auxiliary wall-to-wall spatial predictors and out-of-sample validation; plot-level data must not be disclosed or reverse-engineered. Hourly forest forecasts are not justified by FIA observations alone. Historical hurricane datasets alone do not yield reliable hourly operational forecasts without additional atmospheric forecast inputs.

Weekly Schedule — Spring 2027 (16 Weeks)

Week	Meeting	Expected activity/milestone
1 Jan 6–8	In-Class 1	Kickoff; problem framing, data permissions, computing/GitHub environment; define output scales and probability event.
2 Jan 11–15	Progress 1	Read source guides; download pilot FIA/FHM and hurricane data; document coverage and units.

3 Jan 19–22	Progress 2	Complete raw-data schemas, data dictionaries, source/version manifest; begin Task 1.
4 Jan 25–29	In-Class 2	Data architecture/code audit; reconcile hurricane sources and FIA tables; Task 1 due.
5 Feb 1–5	Progress 3	Implement storm-time/geospatial normalization; construct test spatial units and weather exposure proxies.
6 Feb 8–12	Progress 4	Complete joins, duplicate/ID validation and privacy checks; Task 2 due.
7 Feb 15–19	In-Class 3	Article presentation 1; discuss probabilistic risk and spatial forest modeling.
8 Feb 22–26	Progress 5	Exploratory maps and temporal distributions; missingness/coverage and QA; Task 3 due.
9 Mar 1–5	Progress 6	Develop hurricane event labels, time windows, baselines and local exposure features.
— Mar 8–12	Spring Break	No scheduled meetings.
10 Mar 15–19	In-Class 4	Article presentation 2; one-page modeling plan (targets, predictors, validation and uncertainty).
11 Mar 22–26	Progress 7	Create forest-attribute features and lagged hurricane measures; fit baseline models; Task 4 due.
12 Mar 29– Apr 2	Progress 8	Train and tune candidate models; spatial and temporal holdout evaluation.
13 Apr 5–9	In-Class 5	Research proposal pitch and model review; test one-acre and fine-time claims for scientific validity.
14 Apr 12–16	Progress 9	Calibrate hurricane probabilities; evaluate forest RMSE/MAE and compositional metrics; quantify uncertainty.
15 Apr 19–23	Progress 10	Complete Task 5, prototype interface, maps, sensitivity analyses and peer audit.
16 Apr 26–30	In-Class 6	Final research report, presentation/demo and fully documented GitHub repository.

Weeks follow the Spring 2027 calendar: classes begin Jan 6, there are no classes on Jan 18 (M.L. King Jr. Day) or during Spring Break (Mar 8–12), classes end Apr 23, and Week 16 falls in the final exam period (Apr 26–30). Meeting days and times, task deadlines and presentation assignments will be posted once the class schedule is confirmed. The original V3 assessment structure is retained, but the tasks and readings are tailored to hurricane–forest research.

Suggested Reading and Data Documentation

- USDA Forest Service, FIA DataMart and FIADB user guides: https://research.fs.usda.gov/products/dataandtools/fia-datamart
- USDA Forest Service, Forest Health Monitoring and FIA field guides: https://www.fs.usda.gov/science-technology/forest-health-protection/monitoring
- NOAA NHC, HURDAT2 data and format guidance: https://www.nhc.noaa.gov/data/
- NOAA NCEI, IBTrACS archive and variable metadata: https://www.ncei.noaa.gov/products/international-best-track-archive

- NOAA NCEI, Storm Events bulk CSV documentation: https://www.ncei.noaa.gov/access/storm-events-database/
- Research articles on probabilistic calibration, spatial cross-validation, forest inventory small-area estimation, and hurricane impacts on forests (selected with instructors).

Policies

All relevant Auburn University Student Policy Handbook requirements apply: https://www.auburn.edu/student_info/student_policies/. Students should review the academic honesty, attendance and classroom behavior policies. Accommodations should be coordinated through Auburn University's Office of Accessibility; make-up work will follow university policy and instructor arrangements.

Use of Generative AI Tools

Students may use AI tools (e.g., ChatGPT or Copilot) for conceptual explanations, debugging, documentation and exploratory code suggestions. All submitted analyses, figures and code must be understood, verified, cited as appropriate, and reproducible by the student. Students remain responsible for source integrity, research ethics, privacy and testing of AI-generated output.

Data Security and Research Integrity

Respect all data-use restrictions and confidentiality provisions for forest inventory locations. Report measurement intervals, uncertainty, geography, predictor availability and model scope transparently. Avoid interpreting reported property/crop losses as forest measurements, equating storm maximum winds with local winds, or presenting interpolated hourly/acre-scale forest attributes as direct observations.