

Biographical Sketch – Joseph M. Weaver

Name: Joseph M. Weaver

Position Title: Ph.D. Student, Department of Statistics and Probability, Michigan State University

Email: weave151@msu.edu

PROFESSIONAL QUALIFICATIONS

Michigan State University Statistics and Probability Ph.D. (in progress)

Michigan State University Statistics M.S. (2024)

Central Michigan University Business Administration B.S. (2022)

Lansing Community College Mathematics and Physics A.S. (2013)

ACADEMIC AND PROFESSIONAL POSITIONS

- **Consultant**, Land Core — Data Pipeline and Risk Model Development (2022–Present)
Developed scalable, reproducible data pipelines for agricultural resilience and risk modeling, including ingestion of multi-terabyte geospatial datasets into PostGIS and DuckDB. Designed and optimized ETL workflows using Python, R, and HPC environments (MSU HPCC). Estimated weekly effort: ~10–12 hours/week (2022–2024).
- **Senior AI Engineer**, Jackson National Life, Lansing, MI (2018–2022)
Led AI and data engineering architecture for the enterprise analytics platform. Developed containerized Spark pipelines for datalakes (Cloudera/ADLS) with self-healing and idempotent design. Designed a CRISP-DM-based ML lifecycle integrating compliance, risk, and security requirements.
- **Senior Capacity Planner**, Jackson National Life (2011–2018)
Created automated regression and cluster analyses to forecast infrastructure usage, reducing capacity incidents to near zero and enabling an 80–90% reduction in data center footprint.
- **Senior Software Developer**, Jackson National Life (2005–2011)
Engineered automation frameworks in C# for nationwide broker-dealer systems, replacing 100+ disparate scripts with a unified .NET solution managing 10,000+ production jobs.

Publications and Research Products

- **Weaver, J. M.** (2024). Learning the Intrinsic Dimension of Complex Large Datasets. Master's Thesis, Michigan State University. <https://d.lib.msu.edu/etd/51921>
Abstract: ...bridging manifold learning, statistical geometry, and probabilistic model selection — contributing to the theoretical understanding of intrinsic dimension estimation in non-Euclidean spaces.
- **Pizzo, G., & Weaver, J. M.** (2024). *Crop Yield Risk Bayesian Model (v1.0)*. GitHub Repository.
Implements the large-scale Bayesian hierarchical crop-yield-risk model developed for Pizzo's Ph.D. dissertation; parallelized by Weaver for execution across 10,000+ CPU cores. (GitHub Repository: <https://github.com/landcore/Riskmodel>)
- **Pizzo, G., Weaver, J.M., Viens, F., et al.** (2026+) *Bayesian Prediction of Corn Yield Across Five Midwestern States: Risk Mitigation and Upside Opportunities via the Adoption of More Complex Crop Rotations*. In progress.

- **Weaver, J.M.** (2024). *Land Core Geospatial ETL Toolkit (v1.0)*. (GitHub Repository: <https://github.com/landcore/Riskmodel-ETL>)

Scholarly and Synergistic Activities

- **Bridging Statistical Theory and Computation:** Integrates Bayesian inference, statistical learning theory, and large-scale computational modeling to create reproducible analytical frameworks that connect theory to real-world decision systems.
- **Innovation in Computational Infrastructure:** Designed and deployed distributed data ecosystems across MSU HPC, Docker/Singularity containers, Hadoop clusters, and cloud platforms such as Google BigQuery and Microsoft Azure (Spark-based analytics). Experienced in building data lakes, ETL pipelines, and scalable processing architectures using PostGIS and GitHub CI/CD.
- **Programming and Hardware Integration:** Python, R, SQL, and C# (primary); Java, C/C++, Perl, Bash (secondary). Experienced in low-level systems programming, including EPROM and microcontroller development. Designed and built two fully operational CNC machines using Arduino-based control systems, integrating firmware, electronics, and motion-control software.
- **Applied Theoretical Integration:** Translates advances in learning theory, stochastic processes, and Bayesian modeling into practical architectures for risk prediction, sustainability, and resilience.
- **Reproducible Research Pipelines:** Created YAML-based data dictionaries, GitHub Actions workflows, and automated documentation pipelines to ensure transparency, reproducibility, and continuous integration across large-scale modeling projects.
- **Applied Research Collaboration:** Partnered with Rice University, Land Core, and Michigan State University teams to develop Bayesian crop-risk models and geospatial data systems supporting agricultural sustainability; implemented workflows for 5+ TB of spatial and environmental data.
- **STEM Education and Mentorship:** Instructor for *STT 200: Introduction to Statistics* at Michigan State University; developed and taught programming courses for underrepresented youth through the Boys & Girls Club, emphasizing computational literacy and creative problem-solving.
- **Industry Leadership:** Created and led enterprise-scale AI engineering frameworks at a Fortune 500 company prior to earning a degree—recognized for technical excellence typically reserved for senior architects; experience bridging corporate AI infrastructure and academic research.

Current and Pending Support

Source	Role	Project Title	Dates	Total Award	% Effort
Michigan State University	Graduate Instructor (TA)	STT 200: Introduction to Statistics	2024–Present	N/A (tuition remission + stipend)	~10–15 hrs/week (~20%)
Land Core	Consultant	Soil Resilience and Risk ModelingS	2022–Present	[Lily to fill]	~10–15 hrs/week (~20%)

Collaborators and Affiliations

Viens, Frederi – Rice University (Advisor, Collaborator), Pizzo, Gina – Rice University (Collaborator), Daniel, Lily – Land Core (Collaborator), McLauchlan, Aria – Land Core (Collaborator), Cross, Harley – Land Core (Collaborator), Bhattacharya, Shrijita – Michigan State University (Graduate Advisor)