Matthew Piekenbrock Curriculum Vitae

(937) 269 8582 • matt.piekenbrock@gmail.com • mattpiekenbrock.com

My interests are in unsupervised learning, statistical learning theory, topological data analysis, computational geometry, and building software for scientific computing and reproducible research.

Education

Northeastern University **Boston, MA**
(Pursuing) PhD of Computer Science GPA: 3.83 *2021-Present*

Michigan State University **East Lansing, MI**
(Transferred) PhD of Comp. Mathematics, Science, and Engineering, GPA: 3.50 *2019-2021*

Wright State University **Dayton, OH**
Masters of Science in Computer Science GPA: 3.83 *2015-2018*

Wright State University **Dayton, OH**
Bachelor of Science in Computer Science GPA: 3.42 In-Major, Minor in Statistics *2010-2015*

Teaching Experience

Graduate teaching assistant:

- Northeastern University - Data Mining Techniques (CS 6220 / DS 5230), Summer 2023
- Northeastern University - Supervised Machine Learning (CS 6140/4420), Spring 2023
- Northeastern University - Unsupervised Learning (DS 5230), Fall 2022
- Michigan State University - Computational Modeling & Data Analysis (CMSE 201), Fall 2020

Experience

Graduate Research Assistant **Fall 2021-Present**
Northeastern University *Graduate Student*

My advisor at MSU, Jose Perea, accepted a joint appointment offer transferring to Northeastern University in Boston, MA starting in the Fall of 2021. I transferred with him, continuing my research as a 3rd year graduate student in the Khoury College of Computer Science. During this time I focused primarily on topological dimensionality reduction (see **tallem** in the **Open Source** section) using fiber bundle theory and on spectral-relaxations of the persistent rank invariant.

LERCIP Intern **Summer 2022**
John H. Glenn Research Center at Lewis Field *National Aeronautics and Space Administration*

I was re-hired back at NASA as part of the Space Communications and Navigation (SCaN) program to expand the algorithmic theory necessary to have effective satellite communications in space environments. My research focused on incorporating additional geometric assumptions into routing models built for of delay- and disruption-tolerant networks, particularly in the low Earth orbit regime.

Graduate Research Assistant **Fall 2019-Summer 2021**
Michigan State University *Graduate Student*

I started a PhD program at MSU in Fall 2019, where I spent two years passing qualifying exams and learning the background material necessary to do research in Persistent Homology (PH). My research during this time focused on developing computationally tractable extensions of PH in dynamic and multi-parameter settings, and in showing viable applications of these extensions. This work culminated in an extension to PH that significantly improved the efficiency of the standard reduction algorithm in coarse dynamic settings [4].

Research Associate **Fall 2018-Fall 2019, Fall 2017**
Air Force Research Laboratory *Oak Ridge Institute for Science and Education*

In a collaborative effort to foster new research frontiers in the area of Topology Data Analysis (TDA) between WSU and AFRL, I worked in a research group studying how to combine techniques from the field of topology and machine learning in both supervised and unsupervised settings. I primarily researched multi-scale extensions to the *Mapper* framework, an often used modality for performing TDA. The effort required developing a number of custom open source packages, such as the **Mapper** and **simplextree** packages (see the **Open Source** section).

LERCIP Intern **Summer 2018**

John H. Glenn Research Center at Lewis Field *National Aeronautics and Space Administration*

Towards accelerating materials discovery and design, I was hired by Dr. Steven Arnold (via the Multiscale Modeling Materials and Structures Division) to spend an extended internship at NASA using ML to infer multiscale structural properties from material stress-response data. The project involved deducing process-structure-property (PSP) relationships from a surrogate model trained on laminate stress-strain curve data generated via the Generalized Method of Cells via experimental design theory. My time was primarily spent:

- Learning basic micromechanics and lamination theory
- Architecting a feed-forward neural network (the surrogate model) to model laminate stress-response data
- Implementing a non-parametric information-theoretic estimator efficiently, proving its convergence rates, and modifying an MCMC-like optimization procedure (approximate coordinate exchange) to minimize it

A technical report and subsequent journal publication can be found [6] and [5]. Presentation material, code, and all other material is available upon request for U.S. citizens only.

Graduate Research Assistant **2015 - 2018**

Wright State University *Machine Learning and Complex Systems Lab*

After a brief independent study, I began a graduate research assistantship (GRA) with the Machine Learning and Complex Systems lab studying the use of generative models for modeling macroscopic patterns of real-world traffic networks inferred from raw trajectory (e.g. GPS) data. Topic areas the project focused on included:

- Density-based clustering theory and techniques
- Temporal network models (e.g. stochastic block models)
- Trajectory mining and modeling

Much of my research focused on ensuring the data-inferred networks were representative of the underlying movement data. Our solution involved using the *cluster tree*—a level-set shape characteristic of an estimated [non-parametric] density function—to infer significant clusters of movement [7]. This research was supported by the Center for Surveillance Research, a National Science Foundation I/UCRC.

Student Participant **Summer 2017**

Google Summer of Code 2017 *R Project for Statistical Computing / Google*

I submitted a successful funding proposal under the Google Summer of Code (GSOC) Initiative to the R Project for Statistical Computing to explore, develop, and unify developments related the theory of density-based clustering, namely the recent developments related to the cluster tree. This involved a variety of code development which culminated in the form of an R package, as well as research to further understand the theory and utility of the cluster tree. For more details, see the project page.[1]

Student Research Associate **2014 - 2016**

Air Force Institute of Technology *Oak Ridge Institute for Science and Education*

[1] `https://summerofcode.withgoogle.com/archive/2017/projects/5919718795902976/`

Towards the end of my undergraduate degree, my contract at AFIT was extended under ORISE, where I continued working with the same research group. During this time I primarily worked on the development of a novel Iterative Closest Point algorithm amenable to massive parallelization, implemented in C++/CUDA, for the purposes of enabling real-time tracking of aircraft in the context of Autonomous Aerial Refueling. Our solution involved pairing a cache-oblivious KD-tree search with a novel "Jump-and-Walk" closest-point search on a preprocessed Delaunay triangulation. The effort lead to multiple publications [8, 9]. I also worked on:

- Researching hierarchical markov model for predicting web navigation patterns
- Parallelizing existing atmospheric absorption routines with OpenCL
- Coding a nonlinear optimization algorithm in ANSI-C, and making it callable from MATLAB via MEX

Undergraduate Research Assistant **2013 - 2014**

Air Force Institute of Technology *Southwestern Ohio Council for Higher Education*

I was hired at the Air Force Institute of Technology (AFIT) as an undergraduate student to do research in a multi-disciplinary team called the Low Orbitals Radar and Electromagnetism group, where I worked on a diverse set of projects involving computational, statistical, or physics-based requirements. Being my first research-oriented experience, I either assisted graduate students with primarily programmatic or educational tasks or worked on very computationally-oriented tasks. Some example projects involved:

- Implementing an unsplittable flow approximation algorithm in C++ and Python
- Creating a conversion tool between Oracle's Abstract Data Type and XMLType
- A prototypical UI to to enhance searching and viewing of 2-or-3D models using JavaScript

Publications

Under Development

Matt Piekenbrock Joshua Mike and Jose A. Perea. Tallem: Topological assembly of locally euclidean models. *Available upon request*, 2022.

Matt Piekenbrock and Jose A. Perea. Persistent betti computations in dynamic metric settings. *Available upon request*, 2022.

Journals

Michael Hahsler, **Matt Piekenbrock**, and Derek Doran. dbscan: Fast density-based clustering with R. *Journal of Statistical Software*, 2016.

Conference Papers

Matt Piekenbrock and Jose A. Perea. Move schedules: Fast persistence computations in coarse dynamic settings. *arXiv preprint arXiv:2104.12285*, 2021.

Joshua Stuckner, **Piekenbrock, Matthew**, Steven M Arnold, and Trenton M Ricks. Optimal experimental design with fast neural network surrogate models. *Computational Materials Science*, 200:110747, 2021.

Steven M Arnold, **Piekenbrock, Matthew**, Trenton M Ricks, and Joshua Stuckner. Multiscale analysis of composites using surrogate modeling and information optimal designs. In *AIAA Scitech 2020 Forum*, page 1863, 2020.

Matt Piekenbrock and Derek Doran. Intrinsic point of interest discovery from trajectory data. *arXiv preprint arXiv:1712.05247*, 2017.

Matt Piekenbrock, Jace Robinson, Lee Burchett, Scott Nykl, Brian Woolley, and Andrew Terzuoli. Automated aerial refueling: Parallelized 3d iterative closest point: Subject area: Guidance and control. In *Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016 IEEE National*, pages 188–192. IEEE, 2016.

Jace Robinson, **Matt Piekenbrock** Lee Burchett, Scott Nykl, Brian Woolley, and Andrew Terzuoli. Parallelized iterative closest point for autonomous aerial refueling. In *International Symposium on Visual Computing*, pages 593–602. Springer International Publishing, 2016.

Matthew Maurice, **Matt Piekenbrock**, and Derek Doran. Waminet: An open source library for dynamic geospace analysis using wami. In *Multimedia (ISM), 2015 IEEE International Symposium on*, pages 445–448. IEEE, 2015.

Abstracts

Matt Piekenbrock and Derek Doran. Exploring information-optimal network discretization for dynamic network analysis. *Sunbelt Social Networks Conference of the International Network for Social Network Analysis*, page 262, 2016.

Open Source Contributions

dbscan (R package)[2]	Coauthor
clustertree (R package)[3]	Author
Mapper (R package)[4]	Author
simplextree (R package)[5]	Author
tallem (python package)[6]	Author

Awards, Extra Curricular, Misc.

Outstanding Masters Student Award (Computer Science): WSU 2017-2018 academic year

Student participant and presenter: NSF TRIPODS TGDA Summer School and Workshop

Regional Model United Nations Annual Conference: Served in Volunteer Staff (2016 - 2017)

Outstanding Position Paper Award: National Model United Nations Annual Conference (2014)

Outstanding Delegation Award: National Model United Nations Annual Conference (2013)

[2] https://github.com/mhahsler/dbscan
[3] https://github.com/peekxc/clustertree
[4] https://github.com/peekxc/mapper
[5] https://github.com/peekxc/simplextree
[6] https://github.com/peekxc/tallem