



ERIC M. JOHNSON, PH.D.

eric.johnson643@gmail.com ◇ [Website](#) ◇ [Github](#)

STATEMENT

Expert computational scientist with a proven track record in analyzing complex systems. Passionate educator adept at translating specialized research into accessible knowledge for diverse audiences.

EDUCATION

Ph.D. in Applied Mathematics

Northwestern University, ESAM

Evanston, Illinois

March 2022

M.S. in Applied Mathematics

Northwestern University, ESAM

Evanston, Illinois

May 2016

B.S. *cum laude* in Mathematics and Physics

New York University Abu Dhabi

Abu Dhabi, UAE

June 2014

EXPERIENCE

Mani and Pincus Groups: Postdoctoral Researcher

Northwestern University and The University of Chicago

April 2022 - Present

*Developed novel computational tools for the analysis of scRNA-seq data in *S. cerevisiae*.*

Computational Method Development for scRNA-seq Data

Coordinated across various teams in the Mani and Pincus groups as a computational methods expert. Developed a Pearson residuals-based regression scheme for normalizing scRNA-seq data from yeast. Performed a systematic review of existing scRNA-seq normalization methods to assess their suitability for data generated from yeast. Created user-friendly Python modules to standardize method assessment and implementation. Established and disseminated theoretical heuristics for the assessment of method quality.

Kath Research Group: Doctoral Candidate

Engineering Sciences and Applied Math Dept. at Northwestern University

September 2015 - March 2022

Developed methods and theory for a principled, unsupervised approach to high-dimensional data analysis.

EMBEDR: A Principled Approach to Dimensionality Reduction

Created a computational framework to statistically assess the quality of dimensionality reduction (DR) algorithms. This framework, called **EMBEDR** generates a p -value for each sample in an embedding based on a novel statistical test. Extended EMBEDR into an unsupervised scheme for optimizing user-supplied hyperparameters for common DR methods such as t-SNE and UMAP. Created a novel, parameter-free implementation of the t-SNE algorithm that is reparameterized to be more user-friendly. **Published** a discussion of EMBEDR and results of applying it to atlas-level scRNA-seq data. Published EMBEDR as a **user-friendly Python package** that both executes the algorithm and provides useful visualization tools.

Bayesian Circular-Linear Regression for Prediction of Circadian Phase

Created a Bayesian regression scheme to model circular (angular) variables corresponding to circadian phase measurements as a function of contemporaneously sampled RNA-seq data in human patients in order to investigate the transcriptomic effects of circadian phase disorders and traumatic brain injuries. Developed novel theory to overcome previous barriers in connecting circular variables to inference and regression schemes. Developed novel theory to include regularization to address overfitting and collinearity that arise from the high-dimensionality of RNA-seq data. Proved that this scheme is well-posed and can

provide a unimodal posterior distribution for inferred parameters. Implemented the scheme as a Python module and tested its efficacy on test data.

Multi-Objective Optimization of Models for Clock Neurons in *Drosophila*

Developed a multi-objective optimization scheme for fitting many-parameter models of *Drosophila* clock neurons to *in vivo* measurements of neuronal activity. Engineered informative features from electrophysiology time series data. Developed novel signal processing schemes for noisy electrophysiology time series based on measuring custom robustness metrics. Demonstrated the efficacy of these schemes on real electrophysiology data in a conference presentation. Implemented the feature extraction and model simulation algorithms as a Python package that was used to **link the core circadian clock to other physiological rhythms** in *Drosophila*.

What Do Your Data Say?: Instructor, Creator

Northwestern University

Fall '17, '18, '20; Winter '20; Spring '22

Developed a graduate-level course on introductory statistics and data analysis.

Course Materials using Principles of Backwards Course Design

Published a novel perspective on teaching data science centered on empirical analysis and programming practices. Created course notes, worksheets, classroom activities, lecture materials, assessments, rubrics, and curriculum frameworks using SOTA backwards course design to execute this perspective. Collaborated with education specialists at Northwestern to design and execute surveys to assess the efficacy of course goals. Taught course both remotely and in-person, resulting in over 90% of students meeting goals of being able to perform data analysis tasks such as linear regression, statistical hypothesis tests, and PCA.

An Online Worldwide Bootcamp: What Do Your Data Say?

Adapted the WDYDS course into an online, 4-week bootcamp that enrolled over 600 students around the world. Recreated course materials on an accelerated timeline to educate students and researchers about basic data science during the early months of the COVID-19 pandemic. Created tools, workbooks, and guides to facilitate the instruction of the course by others in an online setting. Created a course website and coordinated teaching assistants to provide students feedback asynchronously. Collaborated with education specialists at Northwestern to design a survey showing that students significantly improved Python programming and data analysis skills.

Python Tutorial for Data Scientists

Created a custom tutorial designed to introduce students with no previous experience to programming in Python. Wrote modules detailing concepts such as looping, parallelization, and object-oriented programming using practical examples. Tested this tutorial with students in WDYDS to assess the tutorial's efficacy in bringing students to a novice level of programming. Published the tutorial as a **website** where students can complete modules using self-directed Jupyter notebooks.

Algorithms Team: Applied Statistics Consultant

Quantum-Si

April 2021 - December 2022

Design and Implementation of statistical algorithms for **Time Domain Sequencing™**

Designed physical models and statistical tests to improve the characterization of fluorescence intensity, binding duration, and binding kinetics distributions for use in downstream algorithms. Was core advisor for subgroup of computational biologists to guide the application of unsupervised and semisupervised methods to discriminate between sequenced amino acids with similar biochemical signatures. Designed custom methods to clean, segment, and classify signals from a massively parallel integrated semiconductor chip in order to annotate amino acids and identify peptide sequences.

The Math Place: Lead Mathematics Tutor

Northwestern University School of Professional Studies

August 2016 - August 2021

Tutored Hundreds of Students in Physics, Math, and Economics

Provided one-on-one tutoring in math and science topics to hundreds of students as head of The Math Place, a free service provided by the School of Professional Studies to Northwestern students. Hired and managed two tutors per year to maintain high standards of student experience. Implemented service-wide tracking of student use in order to improve continuity of student experience across and between academic terms. Designed surveys and performance indicators to provide feedback to faculty and instructors about progress on learning outcomes, trends in student development, and institutional changes in curricula.

Holland Lab: Research Scientist

Courant Institute for Mathematics at New York University

August 2014 - August 2015

Developed and maintained methods and equipment for making meteorological measurements in the field.

Inference of Wind Speed using Ground-Based Imaging

Extended the capabilities of meteorological measurement assays to use inexpensive visual- and infrared-band images by developing novel atmospheric modeling methods. Designed and implemented custom image processing, segmenting, and analysis methods as a Python module. Designed and implemented a custom image deconvolution method to convert wide-angle images into accurate measurements of distance and velocity by reverse-engineering an analytical model of imager lenses and modifying instrument firmware accordingly. Wrote Python programs to automatically process images and send analyses to off-site repositories via a daily satellite link. Developed a novel subsampling method to reduce computational cost while maintaining algorithm accuracy. Methods and accompanying module represented new SOTA for meteorological inference in inhospitable environments (e.g. Antarctic glaciers).

Wind-Driven Upwelling around Grounded Tabular Icebergs

Assessed the impact of a grounded iceberg off the coast of Nova Scotia on the ocean water column by developing custom analysis methods and visualizations of collected oceanographic CTD data. Developed custom regression schemes to quantify the iceberg's effect on the water column. Cross-referenced time series data at different temporal resolutions to demonstrate the presence of wind-driven upwelling only on the upwind sides of the iceberg. Adapted interpolation schemes to generate accurate visualizations of the CTD data angularly around the iceberg. Published the results of this work in the **Journal of Geophysical Research**. The collection of this data was featured in the BBC Documentary *Operation Iceberg*.

Development of Meteorological Data Collection Array

Curated and maintained a rooftop meteorological data collection array at NYU's Courant Institute in order to calibrate and test instruments before deployment in the field. Performed calibration and maintenance of over a dozen instruments, including sonic anemometers, Doppler wind LiDAR, visual- and infrared-band imagers, thermometers, barometers, and rain gauges. Designed and implemented an array-wide data collection scheme to transfer data to an off-site repository daily via a satellite uplink using the CRBasic language. These data and instruments supported the work of researchers at the **Center for Atmosphere Ocean Science** in New York and the **Center for Global Sea Level Change** in Abu Dhabi.

Student Government: Class Representative, Organizing Committee Member

New York University Abu Dhabi

September 2010 - June 2012

Wrote Organizing Documents for the Student Government

As an elected member of the inaugural Student Government Organizing Committee, drafted governing documents outlining the structure of official student representation within NYU Abu Dhabi's operations. Oversaw ratification of this governing document by the student body and organized the first elections of a student government. As class representative, secured over 200,000 AED (55,000 USD) in yearly financial commitment from the university for independent student organizations.

Vinals Group: Undergraduate Research Fellow

University of Minnesota Department of Physics and Astronomy

Summer 2013

Adding Delay into Stochastic Simulation Algorithms for Modeling Genetic Regulatory Networks

In order to model the disparate time-scales involved in the dynamics of gene regulatory networks, developed custom methods based on the stochastic simulation algorithm to include time-delay dynamics. Performed a comprehensive literature review of existing stochastic modeling methods. Implemented custom methods as a Python package and tested their efficacy on toy models. Presented this work as a report and presentation.

AWARDS AND INVITED EVENTS

SciPy Conference 2020 (Workshop cancelled due to pandemic restrictions)

Host, Bringing Data into a 2D World: A Friendly Primer on Dimensionality Reduction

Outstanding Teaching Assistant Award

Northwestern University Department of Engineering Science and Applied Math, 2018

Walter P. Murphy Fellowship

Northwestern University Department of Engineering Science and Applied Math, 2015

James Farley Scholarship

Northwestern University Department of Engineering Science and Applied Math, 2015

Finalist for the Al Khayr Senior Leadership Award

New York University Abu Dhabi, 2014

Full Scholarship Award for Undergraduate Education

New York University Abu Dhabi, 2010

Eagle Scout Rank

Boy Scouts of America, 2010

SKILLS

10+ years of building custom computational tools using a myriad of programming languages and frameworks.

* denotes either direct contributions or reimplementations of a computational method.

Mathematical and Statistical Modeling

Expert in the development of mathematical and statistical models for biological data, with a strong foundation in optimization, information theory, probability theory and linear algebra. Proficient in Maximum Likelihood Estimation, Bayesian Inference, Data Assimilation, and Monte Carlo simulation. Experienced in the use of optimization techniques such as Simulated Annealing, Conjugate-Gradient Methods, and Krylov Methods. Experienced in simulating (stochastic) differential equations using explicit and implicit methods. Experienced in implementing multivariate regression schemes using both Bayesian and ML techniques.

High-Dimensional Data Science

Expert in the responsible use of dimensionality-reduction methods such as PCA*, NMF, t-SNE*, UMAP*, and EMBEDR*. Expert in the design, implementation, and analysis of algorithms for solving complex problems in a variety of data modalities, including Sparse Data, Images, and Time Series.

Programming Languages and Environments

Fluency in Python, R, C++, and MATLAB, with extensive experience utilizing libraries and frameworks such as NumPy, SciPy, TensorFlow, scikit-learn, Bioconductor, and Biopython. Experience with Python package design and development. Significant experience with environments such as Jupyter/IPython, RMarkdown, and Visual Basic Studio. Familiar with IDL, VBA, and CRBasic.

High-Performance and Parallel Computing

Proficient in leveraging high-performance computing (HPC) and parallel processing techniques to handle large-scale biological datasets and improve computational efficiency. Proficient in working with tools and frameworks such as Numba, Cython, MPI, and CUDA, to optimize processing and resource allocation.

Genomics and Bioinformatics

Expert in the design and implementation of algorithms for analysis of several biological data modalities, including RNA-seq, ChIP-seq, single-cell sequencing, fluorescence microscopy, and electrophysiology. Proficient in the use of tools such as DESeq*, Residuals-Based Normalization*, Compositional Data Analysis*, and Persistent Homology. Experienced in using tools such as Seurat, Scanpy, BLAST, Bowtie, SAMtools, Ensembl, and NCBI resources for genome assembly, functional annotation, and comparative genomics.

Systems Biology and Network Analysis

Experience in modeling and analyzing biological networks, including metabolic pathways, protein-protein interactions, and gene regulatory networks. Experience in the use and extension of network analysis methods such as GWAS*, GSEA*, and Diffusion Methods.

Machine Learning and Data Mining

Proficient in applying machine learning and data mining techniques to complex datasets. Experienced in feature selection, model validation, and hyperparameter optimization via methods such as LASSO*, Elastic-Net*, and Cross-Validation to improve model performance. Experienced in implementing and extending unsupervised clustering techniques such as *k*-means*, DBSCAN*, Louvain, and Hierarchical clustering.

Data Visualization and Presentation

Expert in communicating results via data visualization using tools such as ggplot2, Seaborn, and Matplotlib. Skilled in creating clear and concise presentations, reports, and scientific publications to communicate research findings to both technical and non-technical audiences.

Interdisciplinary Collaboration

Adept at working in multidisciplinary teams, liaising with biologists, computer scientists, and statisticians to define project objectives and develop innovative solutions for computational biology research.

Languages: Fluent in **English**, Conversational in **French**

PUBLICATIONS

EMBEDR: distinguishing signal from noise in single-cell omics data.

Patterns, 2022. <https://doi.org/10.1016/j.patter.2022.100443> E.M. Johnson, W.L. Kath, M. Mani.

The E3 Ubiquitin Ligase Adaptor Tango10 Links the Core Circadian Clock to Neuropeptide and Behavioral Rhythms.

Proceedings of the National Academy of Science, 2021. <https://doi.org/10.1073/pnas.2110767118> J. Lee, C. Lim, T.H. Han, T. Andreani, M. Moye, J. Curran, E.M. Johnson, W.L. Kath, C.O. Diekmann, B.C. Lear, R. Allada.

EMBEDR: Separating signal from noise in sc data.

Poster presented at the *UCI Center for Multiscale Cell Fate Research Early Career Researcher Symposium*, April 2021. E.M. Johnson, W.L. Kath, M. Mani.

A general strategy for estimating uncertainty in dimensionality reduction reveals scale and structure in scRNA.

Poster presented at the *Southeast Center for Mathematics and Biology 3rd Annual Symposium*, December 2020. E.M. Johnson, W.L. Kath, M. Mani.

What do your data say? Tools for teaching quantitative approaches in the virtual classroom.

Workshop hosted at the *NSF-Simons Center for Quantitative Biology's Conference on Quantitative Approaches in Biology*, November 2020. E.M. Johnson, C. Freitag, M. Mani.

A feature extraction method for noisy electrophysiology data.

Poster presented at the *Society for Neuroscience Annual Conference*, October 2019. E.M. Johnson, W.L. Kath.

Wind-driven upwelling around grounded tabular icebergs.

Journal of Geophysical Research: Oceans, 2015. <https://doi.org/10.1002/2015JC010805> A.A. Stern, E.M. Johnson, D.M. Holland, T.J. Wagner, P. Wadhams, R. Bates, E.P. Abrahamson, K.W. Nicholls, A. Crawford, J. Gagnon, J.-E. Tremblay.