

Matthew E. Levine

CONTACT INFORMATION	Eric and Wendy Schmidt Center Broad Institute Merkin Building 415 Main St. Cambridge, MA 02142	<i>E-mail:</i> levinema@broadinstitute.org <i>Website:</i> mattlevine.netlify.app
RESEARCH INTERESTS	Machine learning, data assimilation, dynamical systems, Bayesian statistics, bio/physiologic modeling	
EDUCATION	<p>California Institute of Technology, California, CA, USA 2018 — 2023 Ph.D. in Computing + Mathematical Sciences Thesis: <i>Machine Learning & Data Assimilation for Blending Incomplete Models & Noisy Data</i> Adviser: Dr. Andrew Stuart</p> <p>Columbia University, New York, NY, USA 2011 — 2015 B.A. Biophysics, May 2015</p> <p>SIT World Learning Study Abroad, Arica, Chile Fall 2015 Program: Public Health, Traditional Medicine, and Community Empowerment</p>	
AWARDS	<ul style="list-style-type: none">• National Science Foundation Graduate Research Fellowship 2020• Best Paper of the Year on AI in Health 2019 <i>International Medical Informatics Association Yearbook of Medical Informatics</i> Albers <i>et al.</i> "Mechanistic machine learning" JAMIA 2018.• Finalist (1 of 5, Team T2D2), Amazon Alexa Diabetes Challenge 2017• Poster Competition Winner, Data Science Institute, Columbia University 2016• Summer Chemistry Fellow, Société de Chimie Industrielle 2012• National Merit Scholar, Johnson & Johnson Consumer Companies 2011	
RESEARCH EXPERIENCE	<p>Broad Institute, Cambridge, MA, USA</p> <p><i>Postdoctoral Fellow</i>, Eric and Wendy Schmidt Center September 2023 — Present Director: Dr. Caroline Uhler</p> <ul style="list-style-type: none">• Developing mathematical framework for incorporating uncertainty quantification into data-driven learning of model errors in dynamical systems for applications to biomedical timeseries. <p>California Institute of Technology, Pasadena, CA, USA</p> <p><i>Postdoctoral Scholar</i>, Computing + Mathematical Sciences Summer 2023 Adviser: Dr. Andrew Stuart</p> <ul style="list-style-type: none">• Developed mathematical framework for operator-learning with transformer and attention-based architectures (paper forthcoming).• Implemented numerical experiments demonstrating transformers' ability to map between function spaces with resolution invariance (paper forthcoming).• Co-wrote review paper on combining derivative-free optimization methods with machine learning methods to identify structural error models in dynamical systems; targeted towards geophysical science audiences (paper forthcoming). <p>California Institute of Technology, Pasadena, CA, USA</p> <p><i>Graduate Student</i>, Computing + Mathematical Sciences Fall 2018 to Present Adviser: Dr. Andrew Stuart</p>	

- Applied hybrid dynamics modeling framework to learning carbohydrate absorption rates (*NeurIPS Timeseries for Health 2022*).
- Designed physics-based data-driven hybrid-modeling framework for predicting dynamical systems; address markovian and non-markovian model inadequacies in discrete and continuous time (*Communications of the AMS 2022*).
- Implemented novel state-space constraints in an Ensemble Kalman Filter, which performs a constrained state-update via quadratic optimization (*Inverse Problems 2019*).
- Developing modeling and forecasting methodologies for glucose prediction in critically ill patients in Neurological Intensive Care Units and Type 1 Diabetes.

Columbia University, New York, NY, USA

Research Associate, Biomedical Informatics

June 2015 to August 2018

Advisers: Dr. George Hripcsak, Dr. David Albers, Dr. Lena Mamykina

- Implemented and validated non-linear stochastic filtering, Bayesian inverse framework, model averaging, and optimization methods for a *currently deployed personalized blood glucose prediction mobile application* for people with type 2 diabetes (*Plos Computational Biology 2017*).
- Collaborated with social scientists and mobile app developers to ensure successful deployment of the data assimilation technology into mobile applications for use among patients in clinical research studies.
- Developed and evaluated novel methods for time series analysis of non-stationary large-scale electronic health record data, including temporal reparameterizations, linear temporal interpolations, and granger causality (*JBIO 2018*).
- Developed open-source code for evaluating information loss and gain when mapping between medical terminologies, which was shared with an international consortium of medical record stakeholders (totaling over 1 billion patient records) (*JAMIA 2018*).
- Mentored two first-year graduate students in “behavioral phenotyping” projects that used machine learning and self-monitoring data to make personalized recommendations regarding nutrition, sleep, and activity patterns, as well as visually represent these data such that care providers can understand and recognize intra-patient patterns.
- Contributed talks to weekly Data Mining reading group, and presented on classical results from signal processing, machine learning, informatics, and diabetes physiology

TEACHING EXPERIENCE

Teaching Assistant Fellow, Division of Computing and Mathematical Sciences

Caltech Center for Teaching, Learning, and Outreach

AY 2021-2022

- Provide mentoring, training, and support to current teaching assistants in CMS dept.

Teaching Assistant, Data Assimilation and Inverse Problems (ACM 154)

Taught by Prof. Andrew Stuart at Caltech

Fall 2021

- Graduate-level course covering topics in data assimilation (e.g. stochastic filtering, smoothing) and inverse problems (e.g. MCMC, Importance Sampling, Linear Gaussian settings).
- Supported and evaluated students in designing and implementing novel research projects.

Teaching Assistant, Data-driven modeling of dynamical systems (CMS 270)

Taught by Dr. Krithika Manohar at Caltech

Fall 2020

- Graduate special topics course covering state-of-the-art methods in data-driven modeling.
- Supported and evaluated students in designing and implementing novel research projects.
- Delivered guest lecture surveying existing methods for learning model error in ODEs.

Head Teaching Assistant, Machine Learning and Data Mining (CS 155)

Taught by Prof. Yisong Yue at Caltech

Winter 2020

- Graduate/undergraduate course (enrollment of 180 students) covering a broad range of machine learning techniques, with focus on both implementation and theory.
- Coordinated a team of 12 undergraduate TAs to prepare, provide aid for, and grade all homework assignments, projects, and exams.

JOURNAL
PUBLICATIONS

- Melike Sirlanci, Matthew E Levine, Cecilia C Low Wang, David J Albers, and Andrew M Stuart. A simple modeling framework for prediction in the human glucose–insulin system. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(7), 2023
- David Albers, Melike Sirlanci, Matthew Levine, Jan Claassen, Caroline Der Nigoghossian, and George Hripcsak. Interpretable physiological forecasting in the icu using constrained data assimilation and electronic health record data. *Journal of Biomedical Informatics*, 145:104477, 2023
- ME Levine and AM Stuart. A framework for machine learning of model error in dynamical systems. *Communications of the American Mathematical Society*, 2(07):283–344, 2022
- Peter D Sottile, David Albers, Peter E DeWitt, Seth Russell, J N Stroh, David P Kao, Bonnie Adrian, Matthew E Levine, Ryan Mooney, Lenny Larchick, Jean S Kutner, Matthew K Wynia, Jeffrey J Glasheen, and Tellen D Bennett. Real-time electronic health record mortality prediction during the COVID-19 pandemic: A prospective cohort study. *Journal of the American Medical Informatics Association*, 28(11):2354–2365, October 2021
- David J Albers, Paul-Adrien Blancquart, Matthew E Levine, Elnaz Esmaeilzadeh Seylabi, and Andrew M Stuart. Ensemble kalman methods with constraints. *Inverse Problems*, 2019
- David J Albers, Matthew E Levine, Lena Mamykina, and George Hripcsak. The parameter houlahan: a solution to high-throughput identifiability indeterminacy for brutally ill-posed problems. *Mathematical biosciences*, 316:108242, 2019
- Matthew Levine, David Albers, and George Hripcsak. Methodological variations in lagged regression for detecting physiologic drug effects in ehr data. *Journal of Biomedical Informatics*, 86:149–159, 2018
- David J Albers, Matthew E Levine, Andrew Stuart, Lena Mamykina, Bruce Gluckman, and George Hripcsak. Mechanistic machine learning: how data assimilation leverages physiologic knowledge using bayesian inference to forecast the future, infer the present, and phenotype. *Journal of the American Medical Informatics Association*, 25(10):1392–1401, 2018 **Best Paper of the Year on AI in Health, IMIA Yearbook 2019**
- David J Albers, Matthew Levine, Bruce Gluckman, Henry Ginsberg, George Hripcsak, and Lena Mamykina. Personalized glucose forecasting for type 2 diabetes using data assimilation. *PLoS computational biology*, 13(4):e1005232, 2017
- George Hripcsak, Matthew E Levine, Ning Shang, and Patrick B Ryan. Effect of vocabulary mapping for conditions on phenotype cohorts. *Journal of the American Medical Informatics Association*, 25(12):1618–1625, 2018
- Lena Mamykina, Matthew E Levine, Patricia G Davidson, Arlene M Smaldone, Noemie Elhadad, and David J Albers. Data-driven health management: reasoning about personally generated data in diabetes with information technologies. *Journal of the American Medical Informatics Association*, 23(3):526–531, 2016
- Elliot G Mitchell, Esteban G Tabak, Matthew E Levine, Lena Mamykina, and David J Albers. Enabling personalized decision support with patient-generated data and attributable components. *Journal of biomedical informatics*, 113:103639, 2021.

CONFERENCE
PROCEEDINGS &
PRESENTATIONS

- Levine ME. Talk at ICIAM 2023 Minisymposium on Combining Machine Learning and Stochastic Methods for Modeling and Forecasting Complex Systems. [Link to abstract](#).
- K Wang, ME Levine, J Shi, and E Fox. Learning absorption rates in glucose-insulin dynamics from meal covariates. In *NeurIPS 2022 Workshop on Learning from Time Series for Health*, 2022
- Levine ME, Stuart AM. Machine-learning of model error in ODEs. Third Symposium on Machine Learning and Dynamical Systems 2022. The Fields Institute, Toronto, CA. [Link to abstract and recording](#).

- Levine ME. Data Assimilation and Neural ODEs for learning latent dynamics. ICML Workshop on continuous-time methods in machine learning 2022. Poster and paper. [Link to paper](#).
- Levine ME, Stuart AM. Hybrid Data-Driven and Physical Modeling of ODEs in Continuous and Discrete Time. Minisymposium: Model Reduction and Closure without Scale Separation. SIAM Dynamical Systems 2021. Presentation. [Link to abstract](#).
- Levine ME, Stuart AM. Machine-learning of model error in ODEs. Second Symposium on Machine Learning and Dynamical Systems. The Fields Institute, Toronto, CA. [Link to recording](#).
- Levine ME, Albers DJ, Stuart AM, Hripcsak G. Competitive Offline Parameter Estimation for Online Data Assimilation in Glucose Dynamics. Minisymposium: Data Driven Biomedical Dynamics, Modeling, and Data Assimilation. SIAM Dynamical Systems 2017. Presentation. [Link to abstract](#)
- Levine ME, Albers DJ, Hripcsak G. Comparing lagged linear correlation, lagged regression, Granger causality, and vector autoregression for uncovering associations in EHR data. *AMIA Annu Symp Proc*. 2017 Feb 10;2016:779–88. Paper. PMID:28269874
- Levine ME, Mamykina L. Bridging a Gap Between Data Science Research and Health DIY Movement. *Human-Computer Interaction Conference 2016: Advances in DIY Health & Wellbeing Workshop*, May 2016. Workshop position paper.
https://hcihealthcarefieldwork.files.wordpress.com/2015/11/diyhealth2016_paper_5.pdf

INVITED TALKS

- Visiting researcher**, Basque Center for Applied Math, Bilbao, ES **July 2023**
- Seminar Series**, Santa Fe Institute **January 2023**
 Invited Speaker ([Link to recording](#))
- International Ensemble Kalman Filter Workshop**, Norwegian Research Centre (NORCE)
 Invited Speaker **May 2022**
- Johns Hopkins University Data Science Seminar**, Baltimore, MD **April 2022**
 Invited Speaker: "A Framework for Machine Learning of Model Error in Dynamical Systems"

INVITED WORKSHOPS

- Banff International Research Station**, Banff, AB, CA **May 2022**
 Invited Participant
Dynamics and Data Assimilation, Physiology and Bioinformatics
- Mathematisches Forschungsinstitut Oberwolfach**, Oberwolfach, Germany **Feb. 2022**
 Invited Participant, *Data Assimilation: Mathematical Foundation & Applications*
- Berlin Mathematical School**, Zuse Institute, Berlin, Germany **August 2019**
 Participant, *Summer School: The Mathematics of Deep Learning*
- Studied modern theory of neural networks, including their generalizability, expressivity, and applications. Invoked tools from stochastic analysis, dynamical systems, and control to understand neural network properties.
 - Poster: mechRNN—embedding mechanistic dynamics within recurrent neural networks.
- Alan Turing Institute**, London, UK **December 2018**
 Data Study Group Participant, *MedImmune Challenge — Machine learning for enhanced understanding in cell culture bioprocess development*
- Developed supervised time-series forecasting approach for predicting drug production quantity in bioreactors.

North Carolina State University, Raleigh, NC, USA **July 2016**
 Research Training Group in Mathematical Biology
 Participant, *Tutorial Workshop on Parameter Estimation for Biological Models*

- Covered estimation of model parameters and associated uncertainties, parameter identifiability, and uncertainty quantification. Lectures were accompanied by MATLAB exercises.
- Poster: Online state and parameter estimation for personalized, nutrition-based, real-time glucose forecasting in Type 2 Diabetes. Levine ME, Albers DJ, Stuart A, Mamykina L.

Mathematisches Forschungsinstitut Oberwolfach, Oberwolfach, Germany **May 2016**
 Participant, *Data Assimilation Seminar: The Mathematics of Connecting Dynamical Systems to Data*

- Introduction to mathematical and algorithmic foundations of modern data assimilation methods with Bayesian underpinnings.
- Poster: Data Assimilation for Personalized Blood Glucose Forecasting for People with Diabetes. Levine ME, Albers DJ, Stuart A.

ACADEMIC
LEADERSHIP

Minisymposium Organizer **Summer 2023**
 2023 International Conference on Industrial and Applied Mathematics, Tokyo, Japan
 • Co-organized minisymposium on *Randomization for Simplified Machine Learning: Random Features and Reservoir Computers*. [Link to online program](#).

Teaching Assistant Fellow, Division of Computing and Mathematical Sciences
Caltech Center for Teaching, Learning, and Outreach **AY 2021-2022**
 • Provide mentoring, training, and support to current teaching assistants in CMS dept.

SERVICE &
OUTREACH

Caltech Freshman Summer Research Institute, Pasadena, CA **Summer 2020, 2021**
 Graduate Student Mentor
 • Mentored incoming undergraduates interested in machine learning, applied/theoretical math, and programming
 • (2021) Guided two students through exploration of supervised learning, resulting in their own hard-coded implementations of a two-layer neural network and stochastic gradient descent.
 • (2020) Guided the student through an exploration of the logistic map in order to illustrate accessible key principles in programming, applied modeling, and dynamical systems theory.

Mission Unstoppable (TV Show), Los Angeles, CA **2019**
 STEM communication consultant
 • Collaborated with a TV series dedicated to inspiring young girls to pursue STEM careers.
 • Using personal and professional connections to highlight minority voices in STEM.
 • Drafting accessible explanations of fun physical phenomena
 • Collaborate with creatives to develop captivating and accurate content

New York Academy of Sciences, New York, NY **March 2017 to June 2017**
 STEM mentor
 • Led a group of seventh-graders at KIPP STAR College Prep Middle School in a once-a-week afterschool exploration of computer programming using a curriculum adapted from PlayCodeMonkey.org.

Observational Health Data Sciences and Informatics (OHDSI) **2015 — 2016**
 Symposium Planning Committee
 • Supported symposium scheduling, communications, logistics, and funding outreach

Columbia University Office of Residential Programs, New York, NY **2012 — 2015**
 Residential Advisor

- Advise, mentor, mediate, and supervise 63 students on 3 residential floors, providing support for issues regarding mental/physical health, alcohol/drug use, diversity, and academic performance. Create and implement educational and social programs to build community.

Columbia University STRIVE, New York, NY

College Advisor and Mentor

2011 — 2015

- Mentor low-income NYC students with Sickle Cell Anemia in weekly group sessions that provide academic assistance, peer support, and disease management education. Coordinate long and short-term college-prep strategies for 6-12th grade students.

REFERENCES

AVAILABLE TO
CONTACT

Dr. Andrew M. Stuart (e-mail: astuart@caltech.edu; phone: +1-626-395-4560)

- Professor, Computing and Mathematical Sciences, California Institute of Technology

Dr. Emily Fox (e-mail: ebfox@stanford.edu; phone: +1-650-498-3984)

- Professor, Statistics and Computer Science, Stanford University

Dr. George Hripcsak (e-mail: gh13@cumc.columbia.edu; phone: +1-212-305-5334)

- Professor and Chair, Biomedical Informatics, Columbia University

Dr. Tapio Schneider (e-mail: tapio@caltech.edu; phone: +1-626-395-6143)

- Professor, Environmental Science and Engineering, California Institute of Technology

Dr. Yisong Yue (e-mail: yyue@caltech.edu)

- Professor, Computing and Mathematical Sciences, California Institute of Technology

Dr. David J. Albers (e-mail: davidalbers@ucdenver.edu; phone: +1-212-305-5334)

- Associate Professor, Division of Informatics, University of Colorado Medicine

Dr. Lena Mamykina (e-mail: om2196@cumc.columbia.edu; phone: +1-212-305-3923)

- Associate Professor, Biomedical Informatics, Columbia University

JOURNAL REFEREE
PARTICIPATION

- AISTATS
- Physica D: Nonlinear Phenomena
- SIAM Journal on Mathematics of Data Science (SIMODS)
- SIAM Journal on Applied Dynamical Systems (SIADS)
- Nonlinear Processes in Geophysics (NPG)
- Journal of Biomedical Informatics (JBI)
- Mathematical Biosciences
- Mathematical Biosciences and Engineering (MBE)
- Nature: Nutrition and Diabetes