

Christopher M. Jermaine

<http://www.cs.rice.edu/~cmj4>

cmj4@cs.rice.edu

Research Focus **Relational database systems and applied statistical methods.** I focus on developing systems that facilitate the application of methods from probability and statistics to perform analytics over very large data sets. Within this context, my research covers distributed database systems, database query languages and query optimization, applied machine learning and data mining. In addition to developing general-purpose systems software, I work in application areas that include biomedical informatics, genetics, and text analytics.

Education **Georgia Institute of Technology, Atlanta** **June 1998-Dec 2002**
College of Computing, Doctor of Philosophy
Advisor: Edward Omiecinski

The University of California, San Diego **December 1993**
Dept. of Math., Bachelor of Arts, Math/Comp. Sci.

Academic Positions Held **University of Florida, Gainesville** **Fall 2002-December, 2008**
Assistant Professor, CISE Department

Rice University **January 2009-Present**
Associate Professor, Computer Science Department

Notable Systems Built **The SimSQL System.** 120,000 lines of source code. SimSQL is a distributed relational database system augmented to run large-scale recursive machine learning computations (see [9] under Refereed Conference Publications). Web page: <http://www.rice.edu/SimSQL/SimSQL.html>.

The DataPath System. 70,000 lines of source code. DataPath is a high-performance parallel database engine based upon the “data centric” processing paradigm (see [15] under Refereed Conference Publications). This technology has been commercialized by collaborator Alin Dobra (Tera Insights, Inc.)

Awards **Alfred P. Sloan Foundation Research Fellowship**, 2008-2010.

IBM Best Research paper of 2008. For our paper on MCDB.

ACM KDD Conference Best Paper Runner Up, 2009. For our paper on the spatial anomaly detection.

ACM SIGMOD Conference Best Paper Award, 2007. For our paper on the DBO database engine.

National Science Foundation CAREER Award, 2004.

Publications: Summary

Total of nine research papers in SIGMOD, ten in VLDB, six in ICDE. These are regarded as the top data management publication venues. Since 2009: four papers in SIGMOD, four in VLDB, two in ICDE.

Select Refereed Publications

(Student/postdoc authors supervised by Jermaine are noted with an asterisk)

- [1] Risa Myers^{*}, John Frenzel, Joseph Ruiz, Christopher M. Jermaine: Correlating Surgical Vital Sign Quality with 30-Day Outcomes Using Regression on Time-Series Segment Features. In *Proc. SDM 2015*, 11 pages.
- [2] Yanxin Lu, Joe Warren, Christopher M. Jermaine, Swarat Chaudhuri, Scott Rixner: Grading the Graders: Motivating Peer Graders in a MOOC. In *Proc. WWW 2015*, 11 pages.
- [3] Zhuhua Cai^{*}, Luis Leopoldo Perez^{*}, Shangyu Luo^{*}, Jacob Gao^{*}, Zografoula Vagena^{*}, Christopher M. Jermaine: An Experimental Comparison of Platforms for Implementing and Executing Large-Scale Machine Learning Codes. In *Proc. ACM SIGMOD 2014*, 12 pages.
- [4] Luis Leopoldo Perez^{*}, Christopher M. Jermaine: History-Driven Query Optimization with Materialized Intermediate Views. In *Proc. IEEE ICDE, 2014*, 12 pages.
- [5] Zhuhua Cai^{*}, Christopher M. Jermaine, Zografoula Vagena^{*}, Dionysios Logothetis, Luis Leopoldo Perez^{*}: The Pairwise Gaussian Random Field for High-Dimensional Data Imputation. In *Proc. ICDM 2013*: 61, 10 pages.
- [6] Anna Drummond^{*}, Chris Jermaine, Zografoula Vagena^{*}: Topic Models For Feature Selection in Document Clustering. In *Proc. SDM 2013*: 521, 9 pages.
- [7] Supriya Nirkhivale, Alin Dobra, Christopher M. Jermaine: A Sampling Algebra for Aggregate Estimation. *PVLDB 6*(14): 1798-1809 (2013)
- [8] Zhuhua Cai^{*}, Zografoula Vagena^{*}, Luis Leopoldo Perez^{*}, Subramanian Arumugam^{*}, Peter J. Haas, Christopher M. Jermaine: Simulation of Database-valued Markov Chains Using SimSQL. In *Proc. ACM SIGMOD 2013*: 637, 12 pages.
- [9] Zhuhua Cai^{*}, Chris Jermaine: The Latent Community Model for Detecting Sybils in Social Networks. In *Proc. NDSS 2012*, 10 pages.
- [10] Niketan Pansare^{*}, Vinayak R. Borkar, Chris Jermaine, Tyson Condie: Online Aggregation for Large MapReduce Jobs. *PVLDB 4*(11): 1135-1145 (2011)
- [11] Manas Somaiya^{*}, Christopher Jermaine, Sanjay Ranka: Mixture Models for Learning Low-dimensional Roles in High-dimensional Data. In *Proc. ACM KDD 2010*: p 909, 9 pages.
- [12] Fei Xu^{*}, Ravi Jampani^{*}, Mingxi Wu^{*}, Chris Jermaine, Tamer Kahveci: Surrogate ranking for very expensive similarity queries. In *Proc. IEEE ICDE, 2010*: p 848, 12 pages.
- [13] Subi Arumugam^{*}, Alin Dobra, Christopher M. Jermaine, Niketan Pansare^{*}, Luis Leopoldo Perez^{*}: The DataPath system: a data-centric analytic processing engine for large data warehouses. In *Proc. ACM SIGMOD, 2010*: p 519, 12 pages.
- [14] Luis Leopoldo Perez^{*}, Subi Arumugam^{*}, Christopher M. Jermaine: Evaluation of probabilistic threshold queries in MCDB. In *Proc. ACM SIGMOD, 2010*: p 687, 12 pages.
- [15] Peter J. Haas, Christopher M. Jermaine, Subi Arumugam^{*}, Fei Xu^{*}, Luis Leopoldo Perez^{*}, Ravi Jampani^{*}: MCDB-R: Risk Analysis in the Database. *PVLDB 3*(1): 782-793 (2010)
- [16] Mingxi Wu^{*}, Xiuyao Song^{*}, Chris Jermaine, Sanjay Ranka, John Gums: A LRT Framework for Fast Spatial Anomaly Detection. *Proc. ACM KDD 2009*: p 887, 9 pages. **Runner-up Best Paper Award.**
- [17] Alin Dobra, Chris Jermaine, Florin Rusu, Fei Xu: Turbo-Charging Estimate Convergence in DBO. In *Proc. VLDB, 2010* (paper published as *PVLDB 2*(1): 419-430 (2009))