Héctor Corrada Bravo

Postdoctoral Fellow   hcorrada@jhsph.edu
Department of Biostatistics   http://www.biostat.jhsph.edu/~hcorrada/
Bloomberg School of Public Health
Johns Hopkins University
615 North Wolfe St. E5618
Baltimore, MD 21205

Education:

Ph.D., Computer Sciences   August 2008
University of Wisconsin, Madison
Graph-based data analysis
Advisors: Grace Wahba and Raghu Ramakrishnan

Doctor of Musical Arts (ABD)   August 2003
Indiana University School of Music
Guitar Performance and Literature
Bloomington, Indiana

Master of Music   May 1999
Peabody Institute of Music
Johns Hopkins University
Guitar Performance
Baltimore, Maryland

Bachelor of Music   May 1997
Peabody Institute of Music
Johns Hopkins University
Guitar Performance
Baltimore, Maryland

Research Experience:

Postdoctoral Fellow   September 2008-
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Supervisor: Rafael A. Irizarry

> Research in computational genomics, second-generation sequencing,
> gene expression analysis and machine learning

Research Assistant   2003-2008
University of Wisconsin-Madison
Departments of Computer Sciences and Statistics
Advisors: Grace Wahba and Raghu Ramakrishnan

Research in machine learning, bioinformatics, data mining and statistical-relational learning.

Visiting Researcher   Summer 2007
University of Puerto Rico
Departments of Computer Science and Biology
Arbimon Project

Machine learning techniques for automatic species recognition by audio signal processing.

Research Intern   Summer 2004
IBM Almaden Research Center
Advisors: Rakesh Agrawal, Srikant Ramakrishnan and Alexandre Evifimievski

Research in accuracy characteristics of entity resolution algorithms and systems.

Peer-Reviewed Journal Publications:

H. Corrada Bravo, R.A. Irizarry (2009). Model-based quality assessment and base-calling for second-generation sequencing data. *Biometrics,* published online before print November 13, 2009. doi:10.1111/j.1541-0420.2009.01353.x

H. Corrada Bravo, K.E. Lee, B.E.K. Klein, R. Klein, S.K. Iyengar and G. Wahba (2009). Examining the relative influence of familial, genetic and environmental covariate information in flexible risk models. *Proceedings of the National Academy of Science,* 106, no. 20: 8128-8133.

K. H. Eng, **H. Corrada Bravo** and S. Keles (2009). A phylogenetic mixture model for the evolution of gene expression. *Molecular Biology and Evolution.* 26 (10):2363-2372.

M. Acevedo, T.M. Aide, L. J. Villanueva-Rivera, **H. Corrada Bravo,** C. J. Corrada-Bravo (2009). Automated classification of bird and amphibian calls using machine learning: a comparison of methods. *Ecological Informatics.* 4 (4):206-214.

A. T. Evan, R. Bennartz, V. Bennington, **H. Corrada Bravo**, A. K. Heidinger, N. M. Mahowald, C. S. Velde, G. Myhre and J. P. Kossin (2008). Ocean temperature forcing by aerosols across the Atlantic tropical cyclone development region. *Geochem, Geophys. Geosyst.*, 9, Q05V04, doi: 10.1029/2007GC001774.

Peer-Reviewed Conference Proceedings:

H. Corrada Bravo, K. Eng, S. Keles, G. Wahba and S. Wright (2009). Estimating tree-structured covariance matrices via mixed integer programming. *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS '09); Journal of Machine Learning Research Workshop and Conference Proceedings,* 5, 33:40.

H. Corrada Bravo, R. Ramakrishnan (2007). Optimizing MPF queries: decision support and probabilistic inference. *Proc. of the 26th ACM SIGMOD Intl. Conf. on Management of Data* 701:712.

H. Corrada Bravo, D. Page, R. Ramakrishnan, J. Shavlik, V. Santos Costa (2005). A framework for set-oriented computation in inductive logic programming and its application in generalizing inverse entailment. *Proc. of the 15th ILP Conf.* 69:86.

Working Papers:

H. Wu, RA. Irizarry, **H. Corrada Bravo**. "Intensity-level quantile normalization substantially increases mappable reads in ABI's SOLiD sequencing platform." (Submitted)

H. Corrada Bravo, H. Jaffee, B. Langmead, H. Wu, RA. Irizarry. "Fuzzy-Bowtie: a probabilistic pipeline for second-generation sequencing basecalling, quality assurance, mapping and variant discovery."

H. Corrada Bravo, H. Jaffee, B. Langmead, RA. Irizarry. "Probabilstic methods for robust microRNA profiling and discovery using second-generation sequencing."

H. Corrada Bravo, RA. Irizarry. "A note on the impact of technical artifacts on SNP discovery using second-generation sequencing."

H. Corrada Bravo, S. Ding, G. Wahba. "Distance-based regression by regularized kernel estimation."

L. Shan, H.C. Yang, S.A. Rabi, **H. Corrada Bravo,** J.D. Siliciano, R.A. Irizarry, R.F. Siliciano. "Integration of latent HIV-1 in primary resting CD4+ T cells."

A. Rivas, K. Bohane, **H. Corrada Bravo,** M. Tan, R. Tamargo, H.W. Francis. "Facial nerve recovery after vestibular Schwannoma surgery and early predictors of outcome."

Teaching Experience:

2010   "Practical machine learning: methods and algorithmics" principal instructor. Biostatistics Dept., Johns Hopkins University School of Public Health. Graduate level course.

2007   Statistics 860, "Estimation of functions from data". Gave invited lectures. Prof. Grace Wahba. Graduate level course.

2007   Teaching Assistant for CS704 "Principles of Programming Languages". Held office hours for student assistance, graded assignments, created new programming projects. Prof. Thomas Reps. Graduate level course.

2000-2003   Associate Instructor, Indiana University School of Music, Guitar Department. Teach group and individual lessons. Teach undergraduate guitar majors of all levels, from entrance to Senior Recital performance.

Software:

Rcplex: An interface to the CPLEX optimization engine for R. Released Jan. 2008.
http://cran.r-project.org/web/packages/Rcplex/index.html

Rcsdp: An interface to the CSDP semidefinite programming library for R. Released Dec. 2008. http://cran.r-project.org/web/packages/Rcsdp/index.html

Editorial Activities:

Referee for: *Journal of Machine Learning Research, Biostatistics, Journal of Artificial Intelligence Research, Genome Biology, Journal of the Royal Statistical Society (Series C), IEEE Transactions on Computational Biology and Bioinformatics.*

Presentations:

Modeling uncertainty in second-generation sequencing data. *Harvard School of Public Health, Dept. of Biostatistics.* November 2009.

Modeling and managing uncertainty in second-generation sequencing data. *University of Maryland, College Park, Dept. of Computer Science.* October 2009.

Model-based quality assessment and base-calling for second-generation sequencing data. *University of Wisconsin-Milwaukee.* October 2009.

Model-based quality assessment and base-calling for second-generation sequencing data. *University of Manchester.* October 2009.

Model-based quality assessment and base-calling for second-generation sequencing data. *Conference on Next-Generation Sequencing.* Barcelona, October 2009.

Model-based quality assessment and base-calling for second-generation sequencing data. *Third Annual Young Investigators Symposium on Genomics and Bioinformatics.* Baltimore MD, September 2009.

Model-based quality assessment and base-calling for second-generation sequencing data. *WNAR/IMS annual meeting.* Portland, OR, June 2009.

Kernel methods for examining the relative influence of familial, genetic and environmental covariate information in risk models: results and (more importantly) extensions. *University of Wisconsin-Madison,* May 2009.

Model-based quality assessment and base-calling for second-generation sequencing data. *University of Wisconsin-Madison,* May 2009.

Model-based quality assessment and base-calling for second-generation sequencing data. *Case Western Reserve University,* April 2009.

Data analysis at the computational/statistical sciences borderland: two examples from genomics. *Johns Hopkins Bloomberg School of Public Health.* February 2009.

Estimating tree-structured covariance matrices via mixed-integer programming. *Johns Hopkins School of Public Health.* January 2008.

Optimizing MPF queries: decision support and probabilistic inference. *26th ACM SIGMOD Intl. Conf. on Management of Data.* Beijing, China. June 2007.

A framework for set-oriented computation in inductive logic programming and its application in generalizing inverse entailment. *15th ILP Conf.* Bonn, Germany, August 2005.

Posters:

Estimating tree-structured covariance matrices with mixed-integer programming. *12th AISTATS Conference.* Clearwater, FL, April 2009.

Towards and intensity-based error model for short read alignment. *Statistical and Computational Challenges in Next-Generation Sequencing.* Berkeley, CA, October 2008.

Estimating tree-structured covariance matrices via mixed-integer programming with an application to phylogenetic analysis of gene expression. *Second Annual Young Investigators Symposium on Genomics and Bioinformatics.* Baltimore, MD, September 2008.

Honors and Awards:

2003-2007   Ford Fellowship, National Academies of Science.

2004-2005   Advanced Opportunity Fellowship, University of Wisconsin-Madison, Dept. of Computer Sciences.

2001   Second Place, Classic Guitar National Collegiate Artist Competition, Music Teachers National Association.

Skills:

Programming Languages: R, Matlab, C/C++, Python, Java, Prolog, Scheme, ML.
Database Management Systems: PostgresSQL, MySQL.
System Administration: Mac OS X, Linux (Debian/Red Hat).

Other Information:

- Presented concerts as soloist in San Juan, Puerto Rico; Baltimore, Maryland; Washington, District of Columbia; Brasilia, Brazil; Bloomington, Indiana.
- Produced, wrote, and hosted a radio music program for Public Radio in San Juan, Puerto Rico. (1997-1998).