

# Sai Xiao

## Present Address

172 Campbell Dr.  
Mountain View, CA 94043  
(408) 666-7117

## Email

sxiao@soe.ucsc.edu

**Objective** Data Scientist / Quantitative Analyst / Statistician.

**Education** **Ph.D** in Statistics, AMS Department, UC Santa Cruz 2010.09 - 2015.06 (expected)  
Advisor: Athanasios Kottas and Bruno Sansó  
Research topic: Bayesian non-parametric modeling, Point processes.

**M.S.** in Computer Science, Peking University, Beijing, China 2008.07  
Thesis: Design and Implementation of a Decision-Centric Architecture Modeling Tool

**B.E.** in Computer Engineering, Beihang University, Beijing, China 2005.07  
GPA: 3.8/4.0

**Research Experience** **Bayesian non-parametric modeling for point processes.** UC Santa Cruz  
*Thesis Research* 2011.9 - present

- **Marked point processes** We explore a Bayesian non-parametric method to model the time varying intensity of seasonal point processes along with the associated marks. The basic assumption is the points follow a non-homogeneous Poisson process (NHPP). We decompose intensity function as normalizing constant and density function to solve the intractability issue of likelihood of NHPP. The main techniques in the model include using the dependent Dirichlet mixture models and time series models. Various inferences and forecast can be obtained to reveal the characteristics of time-dependent intensity function and conditional density of marks. The work focuses on application in hurricane landfalls along US Gulf and Atlantic coast over 110 years.

*This work won 2014 student paper competition of the Section on Bayesian Statistical Science of the American Statistical Association.*

- **Non-Poisson point processes** We build models to accommodate departures from the Poisson process assumption to tackle more realistically some of the applications considered in this research. We focus on Bayesian non-parametric modeling for renewal process. The model provides flexible shapes of inter-arrival time densities and is capable of modeling point processes with clustering or regularity pattern. Several experiments in synthetic data sets and real data sets are conducted. Another extended model is for continuous-time Markovian renewal processes. The intensity function is influenced by time interval from the last event and the current time as well.

**Research on Architectural Design Decisions**, Software Engineering Institute, Peking Univ., Beijing, China. 2006.09 - 2008.07

- Developed a tool to support decision-centric design method and reuse of design decisions. It helps designers to automatically synthesize the candidate architecture. Using Java in Eclipse.
- All works were collaborated with Dr. Xiaofeng Cui and were published two papers in International conferences.

**Work  
Experience**

**PhD Tech Intern**, eBay Inc., San Jose, CA 2013.06-2013.09

- Worked in search recall team of search science.
- Developed a demo to integrate an attribute extraction tool for search query. Extracted attributes from queries by using conditional random field techniques. Evaluated the precision and recall on top queries in eBay.
- Proposed and implemented a probabilistic model to prune unnecessary query expansions based on extracted attributes.

**Statistician Intern**, Liberty Mutual Insurance, Boston, MA 2012.06-2012.09

- Worked in claim analytics, commercial market.
- Built prototyping predictive models with Generalized Linear Models, Logistic Regression (with regularization), Decision Trees and Boosted Decision Trees.
- Preliminary data processing, variables selection and creation. Provided insight for claim handling process via exploratory data analysis.

**Graduate Student Instructor**, Statistics and Applied Mathematics Department, UC Santa Cruz 2012.03- 2014.03

- Instructor of Biostatistics Lab (AMS 7L). Instructing 150 students weekly on how to use JMP to do fundamental statistical analysis, such as Hypothesis test, ANOVA, linear regression and optimization.

**Software Engineer**, China Foreign Exchange Trade System & National Interbank Funding Center, Shanghai, China. 2008.07-2009.07

- Teammate of the Chinese Yuan Interbank trading system project.
- Responded for requirement eliciting and development of Pricing Engine in derivative market ( using C in Linux platform)
- System testing for the next Chinese Yuan trading system.
- Training assistant for 300 traders on how to use the trading system for one month.

**Publications &  
Working Papers**

**Sai Xiao**, Athanasios Kottas, Bruno Sansó. Modeling For Seasonal Marked Point Processes: An Analysis of Evolving Hurricane Occurrences. To appear in Annals of Applied Statistics.

**Sai Xiao**, Athanasios Kottas, Bruno Sansó. Nonparametric Bayesian modeling and inference for renewal processes. Draft manuscript, in preparation for submission to Technometrics.

Xiaofeng Cui, Yanchun Sun, **Sai Xiao**, Hong Mei. Architecture Design for the Large-Scale Software-Intensive Systems: A Decision-Oriented Approach and the Experience, the 14th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS 2009).

Xiaofeng Cui, Yanchun Sun, **Sai Xiao**, Hong Mei. A Decision-Centric Architecture Design Method Facilitating the Contextually Capture and Reuse of Design Knowledge, the 20th International Conference on Software Engineering and Knowledge Engineering (SEKE 2008).

**Presentations**

Modeling For Seasonal Marked Point Processes: An Analysis of Evolving Hurricane Occurrences. Oral presentation, JSM 2014, Boston, USA.

A Model For Time-Varying Hurricane Intensity. Poster presentation, Graduate Re-

search Symposium, UC Santa Cruz, April 2012.

An Analysis of Evolving Hurricane Wind Speeds and Economic Damages. Poster presentation, Graduate Research Symposium, UC Santa Cruz, April 2013.

**Skills**

- Statistical computing, point process modeling, hierarchical Bayesian modeling, multivariate analysis, variable selection, Markov chain Monte Carlo simulation (e.g. Gibbs sampler, Metropolis-Hasting sampler, etc) and EM algorithm.
- Machine learning methods and algorithms for classification, clustering and regression.
- Proficient with R, C/C++. Familiar and experience with Java, Python, SAS, SQL, JavaScript, HTML.

**CourseWork  
Highlights**

Statistical Inference, Bayesian Statistics, Linear Models, Generalized Linear Models  
Time Series, Stochastic Process, Bayesian Nonparametrics, Multivariate Analysis  
and Machine Learning.

**Honors &  
Membership**

JSM 2014 student travel award	2014
UC Santa Cruz Department of Applied Mathematics and Statistics Teaching Assistantship	2011-2013
UC Regent Fellowship	Fall 2011
Peking University Graduate Scholarship	2005-2008
Tri-A Student in Peking University	2007