Xianlong (Sean) Wang

http://www.seanlog.com

6237 29th Ave NE, Seattle, WA 98115 • xianlong.wang@gmail.com • 541-908-2097

EDUCATION

Doctor of Philosophy, Statistics, Oregon State University, 2009
Dissertation: *Classification for Longitudinal Data*

Master of Science, **Applied Mathematics** & **Computer Science**, Oregon State University, 2008
Thesis: *Chen-Stein Method and Its Applications in DNA Sequence Analysis*

Master of Science, **Statistics**, Oregon State University, 2005
Thesis: *Joint Modeling Clustered Binary & Continuous Responses with Probit Models*

Bachelor of Science, **Material Science & Engineering**, Northeast Forestry University, China, with honors, 1999
Thesis: *Statistical Comparison of Properties of Birch from Different Varieties*

EXPERIENCE

04/09 – Present **Faculty** — Biostat & Biomath program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center

- Develop novel statistical methods for high dimensional data and big data
- Design experiments and conduct statistical data analysis
- Supervise master level statisticians
- Statistical consultant for interdisciplinary colleagues in data analysis, experimental design, and grant application

09/05 – 04/09 **Statistical Consultant** — Statistics Department, Oregon State University

- Design experiments, analyze data, and provide statistical consulting to interdisciplinary projects
- Prepare reports and presentations according to project protocol

12/06 – 04/09 **Research Assistant** — Statistics Department, Oregon State University

- Develop machine learning methods and statistical theory for mining high dimensional longitudinal, including time series and repeated measures
- Present research results to supervisors weekly, and document statistical computer programs
- Mentor undergraduate research assistants in computer simulation, and statistical analysis

06/05 – 09/06 **Computer Administrator** — Statistics Department, Oregon State University

- Maintained and updated the department library SQL database
- Provided services to faculty and graduate students in using statistical software in statistical applications
- Designed the department web pages

09/04 – 04/09 **Teaching Assistant** — Statistics Department, Oregon State University

- Taught lab and recitation for ST541- Probability, Computing and Simulation in Statistics, and ST351- Introduction to Statistical Methods
- Led recitation in ST562, ST563- Statistical Theory and Inferences

03/01 – 09/05 **Research Assistant** — Department of Wood Science & Engineering, Oregon State University

- Designed and implemented Computerized Optimization of Cutting Yield (WinCORY)
- Developed randomized algorithms for forest products automation system computer programs
- Helped implement Computer Aided Lumber Grading (CALG)

PUBLICATIONS

- L K Teixeira, X Wang, Y Li, S E Reed, X Wu, P Wang, S I. Reed. 2015. "Cyclin E Deregulation Promotes Loss of Specific Genomic Regions." *Current Biology*
- B Zhang et al. September 2014. "Proteogenomic characterization of human colon and rectal cancer." *Nature*
- J Hu, X Wang, P Wang. 2014. "Testing Gene-Gene Interactions in Genome Wide Association Studies." *Genetic Epidemiology*
- X Wang, et al. 2013. "A regularized multivariate regression approach for eQTL
- analysis." *Statistics in Biosciences*
- J Kennedy, et al. 2013. "Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins." *Nature Methods*
- X Wang, A Qu. 2013. "Efficient classification for longitudinal data." *ComputationalStatistics and Data Analysis*
- L Chen, J Wang, X Wang, P Wang, 2014. "A mixed effects model for incomplete data with Experiment-level Abundance-Dependent Missing-data Mechanism (EADMM)." *Biometrics* (In review)
- J Gong et al. 2014. "Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer." *Human Molecular Genetics* (In review)

R PACKAGE

groupRemMap
Implements the regularized multivariate regression for identifying master predictors using the GroupRemMap penalty.

PROJECTS

KDD competition, 2008
Develop algorithms for Computer Aided Detection (CAD) of early stage breast cancer from X-ray images.
Apply data mining methods including feature selection, modeling fitting, and cross validation to find the optimal approach to score and classify the image data to malignant and benign cases.

JAK2 as a prognostic biomarker for breast cancer, 2014
An observational clinical study to study the prediction power of JAK2 on breast cancer survival. Calculate the sample size and develop the statistical analysis plan.

Clinical Proteomic Tumor Analysis Consortium (CPTAC)
(http://proteomics.cancer.gov/programs/cptacnetwork)
A comprehensive and coordinated effort to accelerate the understanding of the molecular basis of cancer through the application of robust, quantitative, proteomic technologies and workflows. Lead statistician for statistical method design and data analysis. Apply sophisticated high dimensional statistical methods to high throughput genomic, proteomic, epigenomic data and clinical outcomes.

A Web-Based OSU (Oregon State University) Campus Information System - April 2002
An SQL database system for Oregon State University Campus Facility Department.
Design schema and tables for the relational database, create and test database using MySQL and XML. Document the database and create a help system, present to customers and address requests.

Cross-species (yeast-to-human) comparative systems genetics strategy to identify gene-gene and pathway-pathway interactions underlying human breast cancer, 2012
Use a hybrid, staged approach in which we use an intermediate phenotype (sensitivity to DNA damage) and an unbiased genome-wide search for gene-gene interactions in

Saccharomyces cerevisiae to generate hypotheses that can be tested in GWAS data from humans. Develop statistical methods for integrating information on gene-gene interactions from different sources, and prioritize candidate genetic interactions based on the summarized statistical information. Test interaction candidates using multiple GWAS data sets.

DNA Copy number co-deletion network reveals genetic interactions predictive of breast cancer prognosis, 2010
Use DNA copy number co-deletion network along with integrative genomics analysis to detect genetic interactions that are predictive of breast cancer prognosis.

Improving Nutrient Management with Cover Crops in Organic Broccoli, April 2008
A randomized experiment to determine if the effect of nitrogen fertilizer rates in combination with the selected cover crop treatments for maximum yield of organic broccoli.
Design experiments, formulate the study hypotheses and review study plans. Analyze the data, and interpret analysis results to clients.

Long-term Western Juniper Resource Island Effects Following Cutting, September 2007
An observational study to determine if resource island effects were persistent up to fifteen years following canopy removal. Analyze data with mixed linear model. Interpret analysis results, and prepare project report.

Assessment of Risk Factors for Gestational Diabetes Mellitus (GDM) and Provider Practices for Post GDM Care, September – December 2006
An observational study to answer research questions including to what extent physicians are doing post GDM follow-up care with their patients, what testing criteria Oregon physicians are using to diagnose GDM.
Design study including study hypotheses, questionnaires and endpoints. Calculate sample size, determine study methodology, develop analysis plan including SAS procedures, interpret analysis results.

Gene Expression in Rainbow Trout, *Oncorhynchus mykiss*, Liver During the Stress Response, September – December 2006
A randomized trial to test whether a preselected set of genes are differentially expressed in the livers of rainbow trout, *Oncorhynchus mykiss* responding to an experimental stressor.
Conduct statistical analysis and interpret statistical results for publication.

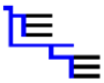
Exercise, Nutrition, and Homocysteine, June - August 2005
An open-label experiment to determine if blood homocysteine levels are different in chronically active individuals when compared to sedentary ones.
Develop trial protocol including formulating study hypotheses, determine endpoints and study populations. Conduct data analysis, and interpret analysis results for submission to publication.

WinCORY (Computerized Optimization of Recoverable Yield for WINDOWS), 2003
A WINDOWS based computer program that calculates random-width, fixed-length cutting yields and best sawing sequences for either rip- or crosscut-first operations.software system.
Design and develop algorithms, implement the algorithms in C++/C. Design the GUI interface.

COMPUTER SOFTWARE/APPLICATIONS

Picture Organizer
An android app for easily organizing and categorizing pictures on android devices. In particular, it allows one to make notes on each picture.

Flex Inventory
An android app for comprehensive and flexible organizing and cataloging daily events, pictures, and other items.



IndexEdtor
Software implemented in C/C++ for organizing, indexing and searching for information.

EnterCeramics
A web application to curate porcelain sales, implemented with HTML, PHP, JavaScript/jQuery, AJAX.

WinCORY
WinCORY (Computerized Optimization of Recoverable Yield for WINDOWS) is a WINDOWS based computer aided design system to help the saw mills in wood industry to optimize the cutting yield.

WinTitle
A WINDOWS application to grab the title of any window so that the title text can be used in any other applications.

EbayCeramics
An web application to browse the porcelain sales on Ebay.

COMPUTER SKILLS

- Proficient in R, JavaScript, PHP, HTML, SQL, MATLAB, C/C++
- SAS Certified Base Programmer for SAS9, December 2007
- SAS Certified Advanced Programmer for SAS9, January 2008
- Proficient in LaTex, MS Office, WINDOWS, Linux

CONFERENCES AND PRESENTATIONS

- Organize invited session “Computational/Statistical Methods for Integrative –omics” at Joint Statistical Meeting, 2015
- TCGA 2nd Symposium, “GroupRemMap - a statistical method for detecting eQTLwith application to TCGA data”, Nov 27-28, 2012, Crystal City, VA
- BioC 2011 - July 28-29, 2011, Fred Hutchinson Cancer Research Center, Seattle, WA
- “A DNA Copy Number Co-deletion Network Identies Genetic Interactions” RECOMB2011, Mar. 28-31, 2011, Vancouver, BC, Canada
- Prentice Symposium, Oct 20-21, 2011, FHCRC, Seattle, WA
- July 31 – August 5, 2010, Classification for Longitudinal Data, Vancouver, BC, Canada
- June 2009, “Classification for longitudinal data”, WNAR, Portland, OR
- April 2009, “Machine learning methods for longitudinal data with application to repeated measurements”, research seminar series, Department of Statistics, Oregon State University

HONORS AND AWARDS

- Jerome C. R. Li Award, Oregon State University, for the top Ph.D. student in Statistics Department, 2008
- "Outstanding Columnist" winner and publication of "Mnemonics Tips" in *B.E.S.T.*, the first English newspaper at Northeast Forestry University, 2000
- "Ten Outstanding Students" winner among 10,000 college students at Northeast Forestry University, 1999
- First Winner of Weinig Scholarship for outstanding senior students, including an $800 prize and a $40,000 planer as a donation to Northeast Forestry University, 1998

PROFESSIONAL MEMBERSHIPS

- The American Statistical Association(ASA)
- Institute of Mathematical Statistics(IMS)