

OBJECTIVE

Senior Data Scientist / Data Center Manager that I can fully utilize my experience in supervising and conducting data fusion, data mining and data visualization processes for enhanced decision support

SUMMARY OF QUALIFICATION

More than 10 years' experience in managing large and complex data sets in various research fields including health care, social science, marketing, climate science and natural resources. Highly skilled programmer in SAS and Fortran. Strong knowledge and experience in statistical applications, business Intelligence systems, database systems, programming languages and mapping tools on UNIX, Linux and Windows. A team worker with strong skill of coordinating between project managers, researchers, students, database administrators and data analysts. Strong organizational skill and multi-task skill. An independent researcher with strong initiative and fast learning ability. Personable, flexible, with high sense of responsibility.

EDUCATION

- **Ph. D.** Electrical Engineering and Atmospheric, Oceanic and Space Sciences, University of Michigan, **2007**
- **M. S.** Atmospheric and Space Sciences, University of Michigan, **2002**
- **M. S.** Atmospheric Physics and Environment, China Academy of Meteorological Science, **2000**
- **B. S.** Atmospheric Physics and Environment, Sun Yat-sen University, Guangzhou, China, **1997**

CONTINUING EDUCATION

- Hadoop Workshop, University of Michigan (Advanced Research Computing), 4/2016
- OpenACC, XSEDE HPC Monthly Workshop, 4/2014
- Applied Survival Analysis, University of Michigan CSCAR, 2014
- Machine Learning, Coursera.org, 2014
- Introduction to Statistical Computing on the Flux Cluster, University of Michigan CSCAR (Center for Statistical Consultation and Research) and ORCI (the Office of Research Cyberinfrastructure), 2013
- High Performance Computing Workshop, University of Michigan ORCI (the Office of Research Cyberinfrastructure), 2013
- SAS Procedures for Analyzing Survey Data, SAS Global Forum Pre-Conference Seminar, 2012
- Applied Structural Equation Modeling, University of Michigan CSCAR, 2012
- Applied Analytics Using SAS Enterprise Miner, SAS Institute, 2011
- Predictive Modeling Using Logistic Regression, SAS Institute, 2010
- Categorical Data Analysis Using Logistic Regression, SAS Institute, 2010
- Forecasting using SAS Software, A Programming Approach, SAS Institute, 2010
- ANOVA and Regression, SAS Institute, 2010
- Microsoft Office PerformancePoint Server 2007 Technical Training, ONLC, 2010
- Introduction to Proc Mixed, University of Michigan CSCAR, 2009
- Using ArcGIS, University of Michigan CSCAR (Center for Statistical Consultation and Research), 2009

- Introduction to Survey Sampling, Summer Institute, Institute for Social Research, University of Michigan, 2009
- **SAS Certified Advanced Programmer**, SAS Institute, 2008
- **SAS Certified Base Programmer**, SAS Institute, 2006
- Introduction to ANOVA, Regression, and Logistic Regression; SAS Macro Language, SAS Institute, 2006
- Probability and Distribution Theory, Department of Biostatistics, University of Michigan, 2005

PROFESSIONAL EXPERIENCE

Lead Programmer Analyst, 4/2015-present

Department of Biostatistics and Medical School (Kidney Epidemiology and Cost Center), University of Michigan, Ann Arbor, MI

- Managing United States Renal Data System (www.USRDS.org)
Supervising the creation and maintenance of USRDS database and reports, including Standard Analysis Files, quarterly update reports, and database for Annual Data Reports
- Supervising programmers on various projects

Senior Programmer Analyst, 8/2013-3/2015

Kidney Epidemiology and Cost Center, Department of Biostatistics, University of Michigan, Ann Arbor, MI

- Produced Dialysis Facility Reports and Quarterly Dialysis Facility Compare reports for ~6000 dialysis facilities across the United States
- Assisted in ESRD database development

Research Area Specialist Intermediate, 5/2008-7/2013

Survey Research Center, University of Michigan, Ann Arbor, MI

Survey Research Center Director's Office 5/2008-7/2009

Statistics and Methodology Unit, Survey Research Operation 7/2009-7/2013

- Developed Survey Research Center Dashboard System to monitor survey projects' daily activity and progress.
 - Administered the whole cycle of the software development, including plan, design, data merge and cleaning, statistical analyses, key performance indicators creation, implementation and maintenance for National Survey of Family Growth (NSFG) Dashboard.
 - Ensured data integrity with well-designed algorithms and data structure.
 - Enabled the development of reports using EXCEL pivot table or OLAP cube in Business Intelligence software with well-designed multi-level data structure.
 - Programs and data structures are with high scalability and can be easily applied to different projects (has been applied to Health and Retirement Study(HRS) and Panel Study of Income Dynamics (PSID)), and can be smoothly run on both Windows and UNIX environment.
 - Automated the daily dashboard update by using innovative methods of scheduling, linking between spreadsheets, and VBA programming.
 - Evaluated SAS Enterprise Business Intelligence and Microsoft Business Intelligence in developing SRC Dashboard System.
- Streamlined workflows
 - Initiated on building efficient computing infrastructures.

- Improved data processing for multiple projects by converting the cumbersome Excel processing procedures into error-proof and efficient SAS programs, and upgraded many old SAS programs into more efficient programs.
- Accelerated data processing by transferring and processing large datasets on Center's Linux clusters. Decreased processing time for large jobs from hours to minutes.
- Advised in data processing, programming and dashboard development for Chinese Family Panel Studies (conducted by Peking University).
- Managed Census data and commercial data
 - Guided decision making of sampling methodologists by producing reports and maps using American Community Survey and Census Decennial data
 - Managed commercial data from Experian, Arbitron, Market Systems Group, Aristotle and US Post Office (Delivery Sequence File). Compared addresses of data from different sources and researched the prediction power of Census data and commercial data in propensity models
- Upgraded predictive models
 - Strengthened consistency between various models by creating call-level and case-level dataset to be shared by multiple modeling projects.
 - Upgraded discrete time hazard model for predicting daily outcome of selected samples using logistic regression modeling.
 - Upgraded propensity model for predicting sample eligibility using logistic regression modeling.
- Sampled for Collaborative Acoustic Project
 - Selected samples for Collaborative Acoustic Project, which include 5000 wave files from 6 surveys by 165 interviewers.
- Analyzed and Documented various projects
 - Conducted quarterly and ad hoc analyses, produced reports for managers and production teams for NSFG (~450K call records, ~50 relational tables, ~300K Audit Trail Data) and HRS (~1 million call records, ~50 relational tables).
 - Produced tables and charts for end-of-cycle data documentation for NSFG.

Sampling Analyst / SAS Programmer, 4/2008

Market Strategies International, Livonia, MI

- Designed/Processed survey samples to measure different customer groups' satisfaction level for different service types provided by utility companies, produced analyses report for project managers.
- Analyzed Henry Ford Health System diabetic patients' web and phone survey samples.
- Analyzed data for utility companies for American Customer Satisfaction Index Project.

Research Fellow, 2007

University of Michigan, Space Physics Research Laboratory, Ann Arbor, MI

- Managed and analyzed field experiment data collected by three research groups; produced report to assist with future data analyses and the design of future field experiments.

Research Assistant, 2000 - 2006

University of Michigan, Department of Electrical Engineering and Computer Science, Department of Atmospheric, Oceanic and Space Sciences, Ann Arbor, MI

- Data Management

- Managed large and complex data set from various sources of satellites, field experiments, GIS, and climate model outputs, with size about 40G on local hard drives. Inquired, processed and stored subsets of terabytes data on HPSS (High Performance Storage System) mass storage, National Snow and Ice Data Center and NASA MSFC satellite data processing center. Data are in various formats of ASCII, binary, spreadsheet, HDF and NetCDF. The platform for most data analyses is UNIX.
- Produced final data products with documentation for field observations. Results are archived in National Snow and Ice Data Center for public access.
- Produced documentation with sample codes for reading and analyzing various model outputs.
- Scientific Programming
 - Adapted various 3-dimensional global or regional climate, hydrology and chemical transportation models for scientific research. Implemented parallel computing algorithms on clusters in Center of Advanced Computing.
 - Supported scientists in Department of Electrical Engineering and Computer Science and Department of Atmospheric, Oceanic and Space Sciences in data analyses/visualization and reporting
 - Developed and coded a new resampling algorithm that is used by NASA as new version of AMSR-E satellite data resampling method. The result is published in IEEE journal.

TECHNICAL PROFICIENCY

- Data & Statistical Analysis
 - SAS, R, MATLAB, EXCEL
- Database Management
 - SQL, Access
- Mapping Tool
 - ArcGIS
- Linux/UNIX Applications
 - Perl, Bash Shell, Emacs, vi, ftp, ssh
- Business Intelligence
 - Microsoft BI (SSMS, SSIS, SSAS, SSRS, PerformancePoint Server2007, Proclarity)
 - SAS BI (Web Report Studio, Information Delivery Portal, BI Dashboard)
 - Pentaho
- Manuscript/Presentation Management
 - Word, Powerpoint, Endnote, RefWorks, Latex, HTML, Photoshop
- Other Programming Languages
 - Fortran, Perl, VBA, C, MPI, OpenMP

PUBLICATIONS

- Gu, Haoyu; Couper, Mick; Kirgis, Nicole; Parker, Sharon; Buageila, Sarrah, "Using Audit Trail Data for Interviewer Data Quality Management", *AAPOR 2013 conference*
- Hubbard, Frost; Wagner, James; Gu, Haoyu, "Building a More Powerful Model to Predict Areas Where USPS-Based Address Lists May Be Used in Place of Traditional Listing", *AAPOR 2013 conference*

- Gu, Haoyu, "Dealing with End-of-Line Markers in Text data shared Across Operating Systems", *SAS Global Forum 2013 conference*
- Hubbard, Frost; Arrieta, Jennifer; Gu, Haley; Guyer, Heidi, "Simple Tools to Determining Best Times to Make Contact for Different Survey Populations", *International Field Directors and Technologies Conference*, 5/2011
- Gu, Haoyu; De Roo, Roger D.; England, Anthony W., "AMSR-E Data Resampling With Near-Circular Synthesized Footprint Shape and Noise/Resolution Tradeoff Study", *IEEE Transactions on Geoscience and Remote Sensing*. 11/2007; DOI:10.1109/TGRS.2007.895412
- Gu, Haoyu, "6.9, 19 and 36 GHz brightness observations of cold lands hydrology in alpine and Arctic terrains", University of Michigan, *Ph.D. Dissertation*, 2007
- De Roo, Roger D.; England, Anthony W.; Gu, Haoyu; Pham, Hanh; Elsaadi, Halim, "Ground-based Radiobrightness Observations of the Active Layer Growth on the North Slope Near Toolik Lake, Alaska", *Geoscience and Remote Sensing Symposium*, 2006. IGARSS 2006. IEEE International Conference.
- Gu, Haoyu; De Roo, Roger D.; England, Anthony W., "The Comparison of AMSR-E Brightness Temperature with Ground-based Observation on the North Slope", *Geoscience and Remote Sensing Symposium*, 2006. IGARSS 2006. IEEE International Conference, DOI:10.1109/IGARSS.2006.452.9/2006.
- Gu, H., H. Pham, Y-C Chung, R.D. DeRoo, and A.W. England, "Model-Based Monitoring of Pan-Arctic Tundra", *ARCSS All Hands Meeting*, Seattle, WA, June 1-3, 2005.
- Gu, Haley, Tzu-yun Lin, Jasmeet Judge, and A.W. England, "A Soil Thermal Conductivity Model Based upon Component Conductivities and Fractions for use in Soil-Vegetation-Atmosphere Transfer Models", *AGU Fall Meeting*, San Francisco, CA, Dec 5-9, 2005.
- Gu, H., and A.W. England, "A Backus-Gilbert resampling scheme for AMSR-E Data that offers improved footprint shape and spatial resolution", *Proceedings IGARSS'04*, Anchorage, Alaska, September 20-24, 2004.
- Gu, Haoyu; Pham, Hanh; Chuang, Yi-ching; De Roo, Roger D.; England, Anthony W., "Model-based monitoring of Pan-Arctic Tundra", *ARCUS 15th Annual Meeting and Arctic Forum 2003*, Arlington, Virginia, April 28-29, 2003.

HORNERS AND MEMBERSHIP

- American Association of Statistics 2012-
- American Association for Public Opinion Research 2012-
- Recommended Graduate Student with Highest Horner (admitted to graduate school in Chinese Academy of Meteorological Science without the entrance exam), Sun Yat-sen University, Guangzhou, Chian, 1997
- ZhangYiLiang Distinguish Student Award First Prize, Sun Yat-sen University, Guangzhou, China, 1996
- Guanghua Distinguish Student Award, Sun Yat-sen University, Guangzhou, China, 1995