# Casson Stallings

## Principal Data Scientist

cassonstallings@gmail.com | 919.264.3692 | Chapel Hill, NC

<https://www.linkedin.com/in/cassonstallings> | github.com/CassonStallings/Publications

Data scientist with a 20-year track record in quantitative and spatial analyses. Proficient in Python, machine learning, and software development. Experience integrating spatial and remotely sensed data with simulation models. Experience contributing to all stages of research and data science projects for diverse clients.

* Developed methodology to increase the classification accuracy of satellite imagery by combining geo-referenced vector data and maximum likelihood spectral classification.
* Automated compound identification in untargeted LC/MS metabolomics data, reducing manual curation time from one or more weeks to one hour for most projects.
* Designed and constructed a 4TB AWS Redshift data base of financial transactions; looked at customer acquisition and growth trends of private companies. Determined basic facts, produced visualizations, and then worked with marketing to generate publications in the financial media.

## Work Experience

**Principal Data Scientist** |Metabolon Mar. 2017 – Oct. 2023

* Designed and coded the Python framework, data pipeline, and some machine learning components for software identifying compounds in untargeted LC/MS metabolomics data. This reduced manual processing time for projects from a week or more to as little as an hour removing a major constraint to the company's growth.
* Analyzed the variability in human curation to better evaluate the automated processing of metabolomics data.
* Provided ad-hoc database queries and analyses to support the business and scientists.
* Maintained, supported, and documented software developed by the Data Science team.

**Senior Data Science Consultant** | Mammoth Data Sep. 2015 – Aug. 2016

* Developed complex Tableau reports for multiple departments at a mortgage lender. Collaborated with company analysts to understand reports generated by hand in spreadsheets. Adapted analyses for Hadoop-based data. Implemented Impala queries, some over 1000 lines of SQL and pulling data from 10 sources.
* Mentored client analysts in use of Tableau, and the Hadoop cluster.

**Data Scientist** | FutureAdvisor May 2014 – Sep. 2015

* Identified significant weaknesses in free customer diversification compared to managed portfolios which through customer education could be used to convert free to premium customers.
* Created a very large AWS RedShift database of financial transactions and used it to generate data stories.

**Lead Mathematician/Statistician** |Alion Science and Technology Feb. 2005 – Sep. 2015

* Spatially segmented urban population based on modeled traffic, road density and bus lines in a manner expected to correlate to air pollution. Assisted clients in comparing these segments with individual health data.
* Designed and coded a multi-pathway Monte Carlo model predicting human exposure to pollutants. Written in SAS and later R, the model combines diverse spatial-temporal data sources and generates population distributions of exposure.
* Statistically characterized the stability of pollutant samples, giving researchers much greater flexibility in collecting and storing samples.
* Technical lead on numerous projects as a government contractor; provided technical direction and facilitated communication and collaboration between clients, partners, and contractors.

## Core Skills

Machine learning, data mining, modeling, exploratory data analysis (EDA), Statistics, Python, Pandas, Dask, Scikit Learn, JupyterLab, Spatial Analysis, ArcGIS, Azure DevOps, GIT, CI, SQL, PostgreSQL, Oracle, MongoDB, written and oral communication, Excel, Mentoring, R, Power BI, User Interfaces, Pachyderm, PySpark, Tableau, Neo4J, Access, Jira, Software Testing

## Education

North Carolina State University

Ph.D., Forestry (Natural Resources)

University of Southern California

M.S., Biological Sciences (Marine Biology)

Portland State University

B.S., Biology

## Publications

Author or co-author on journal articles (20), presentations (40), reports, and manuals (see github.com/CassonStallings/Publications for references)

## Awards

Congressional Medal of Honor for Service in Antarctica

NASA Graduate Student Fellowship, NASA

EPA’s Scientific and Technological Achievement Award, Level III (received 2), US EPA

Patent 11387087: Method for analyzing small molecule components of a complex mixture …

President’s Award for Excellence (received four), ManTech Environmental Technology, Inc.

## Certificates

SAS Advanced Programming for SAS 9 certificate, SAS

Scalable Machine Learning (Spark), BerkleyX, EdX

Introduction to Big Data with Apache Spark, BerkleyX, EdX

R Programming, John Hopkins, Coursera

Hadoop and the Amazon Cloud, Data University

Topics in Artificial Intelligence, Jet Propulsion Laboratory, Pasadena CA