952 NW 60th St
Seattle, WA 98107
peterdavidsmits@gmail.com

November 10, 2020

jobs@ecohealthalliance.org

Dear Hiring Manager,

I am applying for the Research Scientist position at the EcoHealth Alliance. Based on the description, I think that I am a good candidate for this position because of my background in evolutionary biology and ecology as well as my experience with the R language.

Currently, I am a Data Scientist II at Amazon in Advertising. I lead the design and development of multiple data science products using multilevel statistical models to measure the effects of nudges on daily display ad performance. These projects have let me grow my development skill set. I applied tools like Docker to productionize data processing and modeling pipelines on AWS while working with an international Software Development Engineers and Business Intelligence Engineers.

Previously, I worked as a Post-Doctoral Researcher in Biology at the University California – Berkeley were I developed statistical models for describing marine plankton extinction rates over the last 60 million years. This project allowed me to focus on how to communicate uncertainty and out-of-sample prediction error. At Berkeley, I also co-taught a graduate level course on statistical methods in paleontology, macroevolution, and macroecology using the R and Stan languages.

Thank you for considering my application. I am looking forward to hearing more about this position.

Sincerely,

Peter D. Smits

encl: CV, references

Peter David Smits

Seattle, Washington; 609-933-7042
peterdavidsmits@gmail.com

psmits.github.io/
github.com/psmits
linkedin.com/in/pdsmits/

Skills

Languages/Tools: R, Stan, Python, SQL, bash, LaTeX, git, Docker
Packages:

- **R:** brms, rstan, rstanarm, lme4, ggplot2, tidyverse, tidymodels, shiny
- **Python:** pandas, numpy, beautifulsoup, flask

Statistics: Bayesian statistics, multilevel/hierarchical modeling, survival analysis, longitudinal data analysis
Machine Learning: regression, decision trees/CART, random forests, k-means, hierarchical clustering, regularization, feature engineering/selection, dimensionality reduction/PCA, neural networks

Experience

- **Data Scientist II** Seattle, WA
  Amazon – Advertising *January 2020 – present*
  - Developed Recommendation Impact, a data processing and modeling pipeline for measuring the effect of advertising campaign setup suggestions on daily advertising performance using **R**, **Docker**, and **AWS** tools.
  - Designed multilevel/mixed-effects models for describing effectiveness of 9 automated advertising campaign setup suggestions on display advertisement performance, with Maximum Likelihood and Bayesian implementations, using **lme4** and **Stan**.
  - Coordinated with international team to measure and report the quality and value of 24 automated display advertising peformance insights.
- **Data Science Fellow** Seattle, WA
  Insight Data Science *Sept 2019 – December 2019*
  - Designed Copyprism, a web application for generating draft product descriptions from an image.
  - Implemented web scraper for collecting IKEA product descriptions in **python** using **beautifulsoup**.
  - Fine-tuned GPT-2 natural language deep learning model on IKEA catalog in **python** whose generated text was indistinguishable from human text 26% of the time.
  - Deployed Copyprism web application on AWS using **flask**.
- **Postdoctoral Scholar** Berkeley, CA
  University of California – Berkeley *Sept 2017 – July 2019*
  - Designed multilevel discrete-time survival model using **Stan** to predict species extinction which identified species at risk of extinction within 1-million years with an AUC of 0.78.
  - Created multilevel Bayesian time series model in **Stan** for predicting when rare extinction events were likely to be clustered in time based on geological information across multiple databases.
  - Wrote nine lesson short course on analyzing paleontological and macroecological data using **R**, the **tidyverse**, and **brms**, part of which was used as curriculum for graduate-level paleontology course.
- **Graduate Researcher** Chicago, IL
  University of Chicago *Sept 2012 – June 2017*
  - Identified how differences in mammal species ecologies affected their survival rates over the last 65 million years using a multilevel Bayesian survival model implemented in **Stan** and **R** applied to a database of fossil occurrences accessed via web API.
  - Created hidden Markov birth-death model in **Stan** for estimating fossil species observation, origination, and extinction rates over time, where each rate was modeled as its own multilevel regression which incorporated species ecology and environmental context.
  - Mentored and taught graduate and undergraduate students in R, statistics, Stan, and pedagogy.

Projects

- **mathhammr: R** package and **shiny** web application for generating dice rolls for Warhammer 40k 8th Edition and comparing those rolls to a simulated distribution of results.

Education

- **University of Chicago** Ph.D. Evolutionary Biology June 2017
- **Monash University** M.Sc. Biological Sciences Aug 2012
  - Vice-Chancellor's Commendation for Master's Thesis Excellence
- **University of Washington** B.S. Biology June 2010

References

Michael Foote
Professor
Department of Geophysical Sciences
University of Chicago
Phone: 773-702-4320
E-mail: mfoote@uchicago.edu

Kenneth Angielczyk
Associate Curator
Integrative Research Center
Field Museum of Natural History
Phone: 312-665-7639
E-mail: kangeilczyk@fieldmuseum.org

P. David Polly
Professor
Department of Earth and Atmospheric Sciences
Indiana University
Phone: 612-626-9070
E-mail: pdpolly@indiana.edu