Dear Dr. Ross

As a Peace Corps Volunteer in the early 2000s I lived in a transportation hub near the capital of The Gambia, West Africa. Among the few urbanized animals were the fruit bats who mobbed the mango trees. I didn't think about them much, except when I woke up to an explosive "BANG!" when a mishandled fruit crashed onto my tin roof. In retrospect, I realize I resided at a nexus of the world's most important conservation and health issues: animals, people, and transportation. Now, as a biologist and data scientist with 15 years of experience wrangling and analyzing complex datasets, I have the skills to help predict how species interactions and human movements spread infectious disease.

My data science work has involved five primary tasks: 1) **Data cleaning** with **reproducible workflows**, 2) **Summarizing** and **merging** heterogenous data sources, 3) Modeling with **GLMMs**, 4) **Exploring high-dimensional data** with Unsupervised Machine Learning, and 5) **Visualizing** raw data and complex model results.

I have taught myself almost all the techniques necessary to do this work, and focused self-teaching is a key part of my current work as a Teaching Professor. I am an ecologist, but to deliver relevant content to my Computational Biology students I have become proficient in using R for key aspects of bioinformatics, phylogenetics, and computational genomics. Similarly, for two semesters I co-taught a cell biology lab, which required me to dive into dense literature I was not comfortable with and master new lab techniques.

In contrast to learning new techniques, **project management** is the least glamorous -- but one of the most important -- aspects of my job teaching large lecture classes with 75 to 300 students. Organizing and delivering engaging content at this scale requires me to play the simultaneous roles of writer, producer, director, and web-designer. This involves a constant churn of planning, working, and evaluating to render multiple deliverables each week (lectures, assignments, exams, etc.). I have finely honed my ability to set priorities, manage competing tasks, track progress, and organize project materials.

Wrangling messy data sets from long-term studies has also been a frequent part of my academic career. As a first-year graduate student I was instructed to hand-clean a decade-long dataset and I quickly realized the need to do tasks like this computationally. I'm now passionate about reproducibility and an obsessive RMarkdown user and R package builder, with reproducible workflows for almost everything. For example, I recently created a workflow to randomize the order of questions and response choices for exams.

I also frequently use tools such the Tidyverse to summarize individual-level data into community ecology metrics and have worked extensively reconciling and merging ancillary data sources such as remote sensing and weather data. Almost all my analyzes use Frequentist or Bayesian Generalized Linear Mixed Models (GLMMs), and occasionally Generalized Additive Mixed Models (GAMMs). Most of the studies I have worked on have been under powered, so I have learned how to troubleshoot model convergence, use diagnostic tools on the output, and communicate the inferential limits of the results to my co-authors.

Due to the long-term nature and complexity of the data I work with I'm very agile at using ggplot2 to display longitudinal patterns, summarize dense GLMM results and display high-dimensional data. Recently, I made Machine Learning (ML) a major focus of my teaching, lecturing on general ML workflows and having students use Unsupervised ML methods to visualize super-high dimensional population genomics data.

I would love to have the opportunity to learn more about the Research Data Scientist role and discuss how my experience could bring value to your team.

Sincerely,

Nathan Brouwer

Nathan L. Brouwer, PhD

NATHAN L. BROUWER PHD

Computational ecologist specializing in complex datasets

CONTACT INFO

brouwern@gmail.com

github.com/brouwern

517-898-5440

NATHAN L. BROUWER PHD

Data scientist specializing in complex ecological data

I am an ecological data scientist and educator with 15 years of experience extracting insights from messy data and communicating the results. I specialize in using generalized-linear mixed models (GLMMs) to analyze data from long-term observational studies. I have also taught myself key aspects of bioinformatics, phylogenetics, machine learning and population genomics.
Note - All blue text is hyperlinked to relevant online materials.

RESEARCH & PROFESSIONAL EXPERIENCE

present | 2018

Assistant Teaching Professor - General & Quantitative Biology
University of Pittsburgh Dept. of Biological Sciences

- **R programming**: Develop and deliver curriculum for **Computational Biology** course, focusing on machine learning, reproducibility, and high-dimensional data (4 semesters).
- **Data analysis**: Instructor for **Biostatistics** (1 semester) and consultant on stats curriculum for lab classes.
- **Science communication**: Teach scientific writing, non-majors biology, intro biology lecture, Ecology & Evolution Seminar - African Ecology, and labs.

2018 | 2015

Post-doctoral Research Associate - Avian Conservation
National Aviary of Pittsburgh

- **GLMMS**: Analyzed decade-long multisite tropical bird population and community datasets.
- **Data cleaning**: Cleaned and merged diverse datasets of environmental, remote sensing and organism traits data.
- **R packages**: Implemented models of migratory bird populations as reproducible software.
- **Computational ecology**: Developed sensitivity and uncertainty analyses methods for for migration models.

2017 | 2016

Adjunct Professor - Biological Data Analysis
Duquesne Unv. & California Unv. of PA

- **R programming & data science**: Taught grad (Duq., Spring 2017) and undergrad data analyses courses (CalU, Fall 2016 and 2017).

2015 | 2009

Graduate Student
Department of Biological Sciences, University of Pittsburgh

- Graduate Research Fellow (2009-2010, 2014-2015) and NSF Pre-Doctoral Fellow (2010-2013).
- **GLMMs**: Determined appropriate model structures and analyzed data.
- **Data cleaning**: Updated, cleaned and managed data for decade-long plant demographic experiment.
- **Field work**: Designed and conducted research on plant demography.

CONTACT INFO

brouwern@gmail.com
github.com/brouwern
517-898-5440

EDUCATION

B.S., Biology, with Honors 2002.
Seattle Pacific University.

Phd, Biological Sciences - Ecology & Evolution 2015.
University of Pittsburgh.

SKILLS

R packages

GitHub

Open Science

R programming

GLMMs (`nlme`, `lme4`, `glmmTMB`, `brms`)

GAMMs (`mgcv`)

Reproducible reports & pipelines

RMarkdown

Tidyverse

Data visualization with `ggplot2`

Scientific communication

2008 | 2009

Science Writer
Quantum Scientific Publishing
• Wrote and edited customized biology materials for online charter schools.

2007 | 2008

Graduate Student / NSF Pre-Doctoral Fellow
Dept. of Zoology, Michigan State State University
• **Data Cleaning**: Cleaned, organized and analyzed large-scale avian community ecology dataset.

2006 | 2004

Peace Corps Volunteer - Agroforestry Outreach
National Agricultural Research Institute, The Gambia, West Africa
• **Science communication**: Assisted in staff development and conducted outreach on sustainable agriculture.

2004 | 2002

Infecious Disease Research Scientist
University of Washington Department of Allergy & Infectious Disease
• **Lab work**: Conducted experiments on pathogen cell-adhesion proteins.

PUBLICATIONS - SCIENCE EDUCATION

2022

Ecology for All!
Curated compilation of Open-Access ecology resources and new content.
Gowarnis, **Brouwer** et. al.

• **Original content**:
• Chapter 9: Ecology of Populations, section 9.1
• Chapter 10: Population modeling, sections 10.1, 10.2, and 10.4

2021

Computational Biology for All! vs 0.9
An open access book for bioinformatics and computational biology.
Brouwer

2021

A Little Book of R for Bioinformatics vs. 2.0
Open-access bioinformatics primer.
Coghlan (au.) & **Brouwer** (ed., au)

R PACKAGES (AVAILABLE ON GITHUB)

redstart
R implementation and replication of Runge & Marra (2005) *Modeling Seasonal Interactions in the Population Dynamics of Migratory Birds*.
Brouwer *et al.*

combio4all
A repository of functions and data useful for introductory bioinformatics and computational biology.
Brouwer & Coghlan

PASSIONS

Conservation Biology

African Ecosystems

Forest Ecology

Species Interactions

Open Science & Open data

Science Communication

PUBLICATIONS

2019

Direct effects of a non-native invader erode native plant fitness in the forest understory
Journal of Ecology 108:189–198
Bialic-Murphy, **Brouwer** & Kaliz.

- **GLMM**
- Dryad

2019

Stream acidification and reduced availability of pollution-sensitive aquatic insects are associated with dietary shifts in a stream-dependent Neotropical migratory songbird.
PeerJ 6:e5141
Trevelline, Nuttle, Porter, **Brouwer** *et al.*

- **Unsupervised Machine Learning** (NMDS)
- PeerJ Appendix

2018

DNA metabarcoding reveals the importance of aquatic prey subsidies and the structure of dietary niches in a community of breeding riparian songbirds.
Oecologia 187: 85-98
Trevelline, **Brouwer** *et al.*

- **Unsupervised Machine Learning** (NMDS)

2018

Avian community characteristics and demographics reveal how conservation value of regenerating tropical dry forests changes with forest age
PeerJ 6: e5217
Latta, **Brouwer** *et al.*

- **GLMM**, `ggplot`
- GitHub

2017

Long-term monitoring reveals an avian species credit in secondary forest patches of Costa Rica
PeerJ 6: e3539
Latta, **Brouwer** *et al.*

- **GLMM**, `ggplot`
- Harvard Dataverse

2017

Increased photosynthetic performance of an invasive forest herb mediated by deer overabundance
AoB Plants 9: plx011
Heberling, **Brouwer** & Kalisz.

- **GAMM**
- GAMM - GitHub
- AoB Plants

PUBLICATION TOPICS

Demographic models

Longitudinal GLMMs

Species interactions

Tropical birds

Invasive species

2015

Mutualism-disrupting allelopathic invader drives carbon stress and vital rate decline in a forest perennial herb
AoB Plants 7: plv014
Brouwer, Hale & Kalisz.

• **GLMM**, `ggplot2`

2005

Defining the interaction of the *Treponema pallidum* adhesin Tp0751 with laminin
Infection & Immunity. 73: 7485-7494.
Cameron, **Brouwer**, *et al.*

2004

***Treponema pallidum* fibronectin-binding proteins**
Journal of Bacteriology. 186:7019-7022.
Cameron, **Brouwer** *et al.*