

# Nicholas (Nick) Ball

## Geospatial Data Scientist ■ GIS ■ GeoAI

Experienced data scientist now specializing in geospatial, GIS, and GeoAI

- **Versatile data scientist driving product quality, user value and research.** Articulate technical generalist with a background in astrophysics and a perspective of full-stack data science.
- **Synthesize and create analyses to design solutions for end-to-end data science from raw data through to production.** Highly systematic approach.
- **Confident introvert able to clearly communicate with audiences of all types** through presentations, publications, blogs, technical product demos and user documentation. Train, mentor and lead teams.
- **Openminded, grasp both big ideas and detail, highly organized, responsible, straightforward, honest, logical and even-tempered.**

### GEOSPATIAL DATA SCIENCE EXPERIENCE

#### Freelance Data Scientist

San Francisco Bay Area

*Transitioned from full-stack data science to bring my quarter century experience to geospatial data science, following a longstanding interest.*

**Freelance data science projects** encompassing geospatial data science, geographic information systems, and geospatial artificial intelligence.

- **Common programmatic and graphical tools** for geospatial data and analysis, including GDAL, GeoPandas, PostGIS/SQL, QGIS/PyQGIS/ArcGIS, and TorchGeo.

### FULL-STACK DATA SCIENCE EXPERIENCE

#### Senior Data Scientist

HQ in New York / Remote

#### DigitalOcean

Jul 2023 – Jun 2025

*Paperspace was acquired by DigitalOcean (DO), and their expertise in GPUs and AI formed the foundation of DO's successful move into the AI space from 2024 onwards.*

**Continued Paperspace role** as data scientist spanning AI, Content, and Product teams.

- **DO's success in AI required showing that real large language models (LLMs) could be used end-to-end** for pretraining, finetuning, and inference on DO GPU infrastructure that encompassed single GPU, multi-GPU, and multinode.
- **Contributed important proofs that AI could be used in this way** for LLMs up to 30 billion parameters on 10x 8 Nvidia H100 GPU nodes. Comparison to MLPerf benchmarks showed that we were running at state-of-the-art performance.
- **Collaborated on work that improved DO's retrieval-augmented generation (RAG) performance** on real-world data from 50% to over 90%.
- **Showed with our AI team and AMD that their GPUs provided excellent performance-per-cost** on DO infrastructure, resulting in their addition to the DO software stack.

#### Freelance Data Scientist

HQ in New York / Remote

#### Paperspace

Nov 2020 – Jul 2023

*Founded in 2014 under Y Combinator, Paperspace was a cloud computing company providing hardware and software to allow everyone to do data science. By allowing easy access to infrastructure-as-a-service computing, GPU hardware acceleration, software-as-a-service Notebooks, and MLOps, via its Core and Gradient products, barriers to entry for doing data science were removed for a global audience.*

**Serving as data scientist for the company, spanning the Content, Product, and Research teams.** Working with the CTO, COO, and team leads to guide the product to meet the needs of both beginners and advanced users.

- **Added Data Science space to company internal organization**, initiating and leading 15 projects.
- **Developed advanced data science content on Core and Gradient**, including deep learning recommender systems, gradient-boosted decision trees, TensorFlow Distribution Strategies, Hugging Face Accelerate, PyTorch Lightning, Graphcore IPUs, generalist models (image+text+audio+video), Stable Diffusion with self-attention guidance, and Mosaic MPT & Llama 2 LLM finetuning.

415.409.8390

San Francisco Bay Area, CA 94030

[linkedin.com/in/nickballdatascientist](https://www.linkedin.com/in/nickballdatascientist)

[www.nickballdatascience.com](https://www.nickballdatascience.com)

[github.com/nmball42](https://github.com/nmball42)

[nick@nickballdatascience.com](mailto:nick@nickballdatascience.com)

### TOOLBOX

**Geospatial Data Science, GIS, and GeoAI**

**Data Science End-to-End (full stack)**

**Machine Learning & Artificial Intelligence**

**Customer Data Science Projects**

**Academic Research for Publications**

**Public Speaking, Publications and Communications**

### EDUCATION

#### Ph.D. Astronomy

University of Sussex, UK  
2004

*Studied the large-scale structure of the universe. Thesis: Galaxy Types, Luminosity Functions & Environment in the Sloan Digital Sky Survey*

#### M.Sc. Astronomy

University of Sussex, UK  
Distinction | 2001

#### B.A. Natural Sciences (Geological Sciences)

Cambridge University, UK  
2000

*Included field geological mapping in the Rocky Mountains of Montana with refereed publication*

### GEOSPATIAL DATA SCIENCE EXPERTISE

GDAL / OGR

GeoAI

GeoPandas

PostGIS / SQL

QGIS / PyQGIS / ArcGIS

TorchGeo

### FULL-STACK DATA SCIENCE EXPERTISE

*Deep learning and neural networks*

*End-to-end: data preparation, model training, inference*

*Feature engineering for data*

*Large datasets*

- **Showcased first-on-Paperspace results**, including end-to-end data science, Gradient Workflows in production, and machine learning model Deployments.
- **Produced products available to users:** ML-in-a-Box VM on Core, and Gradient Public Datasets.

## Principal Data Scientist (Product)

HQ in UK / Fully Remote Team

## Dotscience

Sep 2019 – May 2020

*Dotscience was founded and funded in 2017 by DDN, the world's largest privately held storage company, to build an end-to-end data science platform using MLOps (principles of software DevOps for machine learning) for faster deployment and to hold AI accountable for its decision making. (Product launched July 2019; company closed May 2020 due to COVID-19.)*

**Served as de facto chief data scientist and head of product**, working with CEO and VP Engineering to shape the development of the Data Science platform with underlying MLOps architecture to deliver specific value to data scientist users.

- **Provided strategic input to product direction, developed user documentation and tutorials, prepared product demonstrations, and wrote and presented on product capabilities.** Served as the company go-to and manager for data science and machine learning.
- **Owned and drove company roadmap to focus on MLOps**, creating and prioritizing 80+ roadmap items across all aspects of data science.
- **Conducted extensive product testing from advanced user point of view. Gathered from and shared feedback with product**, engineering, design (UX), sales, marketing, and the CEO.

## Lead Data Scientist

Redwood Shores, CA

## Oracle Corporation

Oct 2018 – May 2019

*Database giant Oracle built their own cloud infrastructure (OCI) and acquired datascience.com as part of building the new Oracle data science platform to be launched in 2019. (Part of large-scale layoff.)*

**Joined group led by former datascience.com CEO to become one of two data scientists working on the product team.** Worked closely with the ODSS (Oracle Data Science Service) engineering team, bridging the outward-facing (customer) and inward-facing (product and engineering) aspects of the group's data science work.

- **Learned the ODSS product and OCI environment** and worked with other Oracle teams on product directions on algorithms, the data science process, product integrations, and research.
- **Produced demo content on fraud detection and a talk on data science pitfalls.** Supported customer conversations and visits. Staffed product booth at annual Oracle conference.

## Principal Data Scientist

Palo Alto, CA

## Infosys

Mar 2017 – Oct 2018

*Infosys provides IT services, primarily in India, and wanted to add AI products to its service portfolio for its global customer base. It acquired Skytree's technology and team in 2017.*

**Joined with the acquisition of Skytree, staying with the team** to continue working on the world-class machine learning product, with the potential for it to become global. Served as technical data science representative with salespeople and engineers, occasionally leading calls.

- **Led PoC (Proof of Concept) project for a bank for fraud detection** using supervised machine learning to predict patterns for fraudulent transactions.
- **Project lead for the group's data science work, developing product roadmap and generating Jira tickets** and epics for engineers. Led testing and feedback from data scientist point of view.

## Staff Data Scientist

San Jose, CA

## Skytree

Nov 2012 – Mar 2017

*Skytree was founded in 2010 as a platform to do machine learning, based on the work of the FASTLab group at Georgia Tech. The company grew from 20 to 50 people in my time there. (Acquired by Infosys.)*

**Recruited into a world-class group of data scientists by the CEO after being one of Skytree's first customers** while at the Herzberg Institute. Combined research responsibilities and work on PoC projects and product roadmaps, feedback and demos.

- **Led Skytree benchmarking project with data science and engineering** to prove that Skytree was more than 100x faster than competitors. Results used by company marketing until 2017.
- **Served as lead data scientist on 20+ projects including 5 of Skytree's 10 paying customers**, 10 other PoCs and pre-PoC projects. Generated product feedback totaling hundreds of Jira tickets (roadmap epics, technical, user experience, presentation), significantly improving the product.
- **Authored videos and blogs for Skytree's website**, including technical differentiators, how to download and install, speed and automation, and demonstrations of machine learning across topics. Outlined, scripted, and performed live demos. Represented Skytree solo at trade show.

Large language models (LLMs)

PyTorch

Supervised machine learning:  
gradient boosted decision trees, k-nearest neighbor, random forest, support vector machine

Unsupervised machine learning:  
clustering, outliers

## COMPUTING TECHNOLOGY

C programming language

Cloud computing

Hardware acceleration (GPUs)

LaTeX typesetting

MATLAB

Microsoft Office (Excel, PowerPoint, Word) and equivalents

Python

Shell scripting (Bash)

UNIX/Linux

## ACADEMIC RESEARCH

Astronomy & astrophysics

Astroinformatics

Geology

Scientific writing & publications

## TRANSFERABLE SKILLS

Collaborative and cross-functional problem solving

Communication to technical and nontechnical audiences (presentations, reports, speaking, writing)

Mentoring, teamwork, training

Project management

Technical writing

Touch typing

## ON THE SIDE

Toastmasters International  
Competent Communicator

World Scrabble Championship team member for UK, USA, and Canada. Peak rankings include top 50 world and top 10 in the UK. Past certified tournament director.

Global perspective: UK & US citizen, spouse is Taiwanese, children are American

# Nick Ball Addendum

## ACADEMIC RESEARCH EXPERIENCE

### Assistant Research Officer

Victoria, BC, Canada

Herzberg Institute of Astrophysics

Jul 2009 – Jul 2012

*Herzberg Institute of Astrophysics is the astronomy institute of Canada's National Research Council.*

**Hired for the CANFAR project, a precursor to cloud computing, but for astronomy**, including the Next Generation Virgo Cluster Survey (NGVS). This was a newly created position, uniquely overlapping astronomy and data science, focused on data science for publication.

- **Authored "Data Mining and Machine Learning in Astronomy,"** one of the first major reviews in the world on this topic (invited review, International Journal of Modern Physics D, 61 pages).
- **Created one of the first large-scale astronomy data mining systems in the world**, via collaboration with Skytree: CANFAR+Skytree.
- **Led the Luminosity Function Science Working Group** of the NGVS and the Photometric Redshifts subgroup. Showed that machine learning is able to resolve which galaxies are in the Virgo Cluster.

### Postdoctoral / Research Scientist

Urbana, IL

University of Illinois at Urbana-Champaign

Nov 2004 – Jun 2009

*UIUC's astronomy department & National Center for Supercomputing Applications (NCSA) were part of the nascent astronomy data mining community.*

**As a postdoc, directed research and publications** and served as Principal and Co-Investigator on funding grants including from NASA.

- **Produced 3 peer-reviewed publications of original research in the *Astrophysical Journal***, including the first sample of quasars with accurate distances from just images. Created some of the first applications of machine learning to astronomical datasets of over 100 million objects, using NCSA supercomputers.

## PUBLICATIONS AND PATENTS: SUMMARY

- Lead author, 6 publications in major peer-reviewed journals (Monthly Notices of the Royal Astronomical Society; Astrophysical Journal)
- Lead or co-author, 15 other publications, beginning aged 20
- Total citations 2137; 1351 as first author; h-index = 13 (Google Scholar, February 2026)
- Coauthor of 2 patents for Skytree

## SELECTED ACADEMIC PUBLICATIONS: DETAILS

**Ball N.M.**, 2013, *CANFAR+Skytree: A Cloud Computing and Data Mining System for Astronomy*. Astronomical Data Analysis Software & Systems (ADASS) XXII, ASP Conference Proceedings, eds. Friedel D., Freemon M., Plante R. (San Francisco: ASP)

**Ball N.M.**, 2012, *Utilizing Astroinformatics to Maximize the Science Return of the Next Generation Virgo Cluster Survey*. *Astrostatistics and Data Mining in Large Astronomical Databases*, Springer

Ferrarese L.F., et al., 2012, *The Next Generation Virgo Cluster Survey (NGVS). I. Introduction to the Survey*, *ApJS* 200 4

Gaudet S., Hill N., the CAD team, 2010, *CANFAR: the Canadian Advanced Network for Astronomical Research*. *Proc. SPIE, Software and Cyberinfrastructure in Astronomy*, eds. Radziwill N.M. and Bridger A., 7740-1L

**Ball N.M.** & Brunner R.J., 2010, *Data Mining and Machine Learning in Astronomy*, *Intl. Journal of Modern Physics D*, 19(7), 1049

Myers A.D., White M. and **Ball N.M.**, 2009, *Incorporating photometric redshift probability density information into real-space clustering measurements*, *MNRAS* 399 2279

**Ball N.M.**, et al., 2008, *Robust Machine Learning Applied to Astronomical Datasets III: Probabilistic Photometric Redshifts for Galaxies and Quasars in the SDSS and GALEX*, *ApJ* 683 12

**Ball N.M.**, Loveday J. & Brunner R.J., 2008, *Galaxy Colour, Morphology, and Environment in the Sloan Digital Sky Survey*, *MNRAS* 383 907

**Ball N.M.**, et al., 2007, *Robust Machine Learning Applied to Astronomical Datasets II: Quantifying Photometric Redshifts for Quasars Using Instance-Based Learning*, *ApJ* 663 774

**Ball N.M.**, et al., 2006, *Bivariate Galaxy Luminosity Functions in the Sloan Digital Sky Survey*, *MNRAS* 373 845

**Ball N.M.**, et al., 2006, *Robust Machine Learning Applied to Astronomical Datasets I: Star-Galaxy Classification of the SDSS DR3 Using Decision Trees*, *ApJ* 650 497

**Ball N.M.**, et al., 2004, *Galaxy Types in the Sloan Digital Sky Survey Using Artificial Neural Networks*, *MNRAS* 348 1038

Moore J.R., Watts M.D., **Ball N.M.**, Ratcliff M.D., 2000, *The Geology of the Ford Creek Area, Northern Montana Thrust Belt. Montana/Alberta thrust belt and adjacent foreland*. 50th Anniversary Symposium of the Montana Geological Society, Volume 1, pp 151–168, Montana Geological Society, Billings, Montana, USA