

SARAH C. SHI

sarah.c.shi@gmail.com | github.com/sarahshi | +1 (646) 734-6388

EDUCATION

- 2024- UNIVERSITY OF CALIFORNIA, BERKELEY, DOCTOR OF PHILOSOPHY IN EARTH SCIENCE
Supervisor: Professor Penny Wieser
- 2021-2022 UNIVERSITY OF CAMBRIDGE, MASTER OF PHILOSOPHY IN EARTH SCIENCE
Thesis: Development of Machine Learning Methodology for Automating Petrography and Studying Icelandic Xenoliths
Supervisor: Professor John MacLennan
- 2016-2020 COLUMBIA UNIVERSITY, BACHELOR OF ARTS IN EARTH SCIENCE, *summa cum laude*
Thesis: Run-Up and Syn-Eruptive Dynamics of Volcán de Fuego's Eruption of 2018
Supervisor: Professor Terry Plank

FELLOWSHIPS AND AWARDS

- 2025 HEARTS TO HUMANITY ETERNAL (H2H8) GRADUATE RESEARCH GRANT TO ADVANCE HUMANITY
· Awarded grant (\$10,000) to examine Cascade volcanism, with a focus on advancing humanity
- 2025 NORTHERN CALIFORNIA GEOLOGICAL SOCIETY, RICHARD CHAMBERS MEMORIAL SCHOLARSHIP
· Awarded scholarship (\$2,000) to examine Cascade volcanism in Northern California
- 2024-2026 CHANCELLOR'S FELLOWSHIP, UNIVERSITY OF CALIFORNIA, BERKELEY
· Awarded fellowship recognizing scholarship and academic contributions in the top admitted doctoral students applying to the Math and Physical Sciences Division at University of California, Berkeley; granted full tuition and fees
- 2024 GATES CAMBRIDGE SCHOLARSHIP AND CAMBRIDGE INTERNATIONAL SCHOLARSHIP (DECLINED)
· Granted full scholarship awarded to ~1% of applicants to pursue graduate study given 1) outstanding intellectual ability 2) reasons for choice of course 3) a commitment to improving the lives of others and 4) leadership potential
- 2024 GORDON RESEARCH SEMINAR AND CONFERENCE ON GEOCHEMISTRY OF MINERAL DEPOSITS
· Granted \$300 as invited Gordon Research Seminar speaker, \$500 for the Gordon Research Conference
- 2021-2022 EURETTA J. KELLETT FELLOWSHIP
· Granted fellowship given academic distinction and potential for community contribution at the University of Cambridge; awarded full tuition and maintenance (£45,120) to pursue one year of graduate study
- 2022 EMMANUEL COLLEGE, UNIVERSITY OF CAMBRIDGE — POZZI FUND
· Examined intersection of science and art, sonified thermodynamic data from magma crystallization (£250)
- 2020 COLUMBIA UNIVERSITY
· Phi Beta Kappa; Departmental Honors, awarded for academic and research excellence

RESEARCH

- 2022-2024 DATA SCIENCE FELLOW IN GEOINFORMATICS, LAMONT-DOHERTY EARTH OBSERVATORY
Data Science Fellow Research
· Develop [mineralML](#), an open-source Python applying supervised and unsupervised machine learning (ML) solutions for probabilistic mineral classification in large geochemical databases to ensure data quality through the cascade
· Spearhead integration of data repository APIs of mineral, melt, and astromaterial data with Python, resulting in increased user interactivity with cloud computing solutions and Jupyter notebooks
- 2021-2022 UNIVERSITY OF CAMBRIDGE
Master of Philosophy Research
· Develop and deploy ML solutions for classifying electron microscopy data ($\sim 10^7$ pixels \times 10 chemical channels) with dimensionality reduction and clustering to automate mineral segmentation resulting in the significant diminution of manual intervention (\sim tens-hundreds of hours to \sim 5 hours for each sample); create Bayesian MCMC algorithm to determining mineral chemistries, accounting for covariance in analytical uncertainties
· Model plagioclase chemical profiles with finite-elements diffusion modeling to determine magma mixing timescales; evaluate spatiotemporal variability in Icelandic plagioclase
- 2019-2021 LAMONT-DOHERTY EARTH OBSERVATORY OF COLUMBIA UNIVERSITY
Computational Research Assistant (5/2020-7/2021); **Senior Thesis Research** (3/2019-5/2020)
· Identified the eruption trigger of fresh magma injection, occurring two weeks prior to outset, for the Volcán de Fuego eruption of 2018 by timing magma mixing with finite-difference diffusion modeling
· Developed [PyIRoGlass](#), an open-source Bayesian MCMC Python package to fit baselines to FTIR spectra for volatile phase (CO₂, H₂O) measurements within basaltic-rhyolitic glasses, creating the first standardized data processing method
· Invented novel arc melt thermometer with inverse theory to assess melt temperatures and uncertainties, significantly reducing degrees of freedom and improving accuracy and precision of reconstructed temperature
- 2018-2019 NSF REU Summer Intern (6/2018-9/2018); Independent Research (9/2018-1/2019)
· Analyzed and modeled *n*-alkane distributions and concentrations with dimensionality reduction techniques to reveal resolving power among plant functional types, in application to reconstructing hominid ecosystems

PAPERS

- 2025 [6, in preparation] **Shi, S.C.**, Wieser, P.E., Toth, N., Antoshechkina, P., Lehnert, K., mineralML: Leveraging Machine Learning for Probabilistic Mineral Classification in Geochemical Databases. *JGR: Machine Learning and Computation*.
[5, in review] Toth, N., **Shi, S.C.**, MacLennan, J., Tung, P.Y., EDS Analysis for Petrology: A Probabilistic Framework. *JGR: Machine Learning and Computation*.
[4, in review] Wieser, P.E., **Shi, S.C.**, Gleeson, M., Rangel, B., DeVitre, C., Bearden, A., Lynn, K., Trusdell, F., Camille-Caumon, M., Fluid inclusion constraints on the geometry of the magmatic plumbing system beneath Mauna Loa: Part 1: Extrusive products. *Bulletin of Volcanology*.
[3] Gleeson, M., Wieser, P., deVitre, C., **Shi, S.C.**, Millet, M.-A., Muir, D., Stock, M., Lissenberg, J., Persistent high-pressure storage beneath a near-ridge ocean island volcano (Isla Floraena, Galapagos). *Journal of Petrology*.
[2] Moussallam, Y., Towbin, W.H., Plank, T.A., Bureau, H., Khodja, H., Guan, Y., Ma, C., Baker, M., Stolper, E.M., Naab, F.U., Monteleone, B.D., Gaetani, G.A., Shimizu, K., Ushikubo, T., Lee, H., Ding, S., **Shi, S.C.**, Rose-Koga, E.F., ND70 Series Basaltic Glass Reference Materials for Volatile Element (H₂O, CO₂, S, Cl, F) Analyses and the C Ionization Efficiency Suppressing Effect of Water in Silicate Glasses. *Geostandards and Geoanalytical Research*.
[1] **Shi, S.C.**, Towbin, W.H., Plank, T.A., Barth, A.C., Rasmussen, D., Moussallam, Y., Lee, H., Menke, W., PyIRoGlass: An open-source, Bayesian MCMC algorithm for fitting baselines to FTIR spectra of basaltic-andesitic glasses. *Volcanica*.

CONFERENCES

- 2025 [16] Moussallam, Y., Towbin, H., Plank, T., Bureau, H., Khodja, H., Guan, Y., Baker, M.B., Stolper, E., Naab, F., Monteleone, B.D., Gaetani, G., Shimizu, K., Lee, H., Ushikubo, T., Ding, S., **Shi, S.C.**, Rose-Koga, E.F., Development of Basaltic Glass Reference Materials for Volatile Element Analysis (H₂O, CO₂, S, Cl, F) and Investigation of Water-Induced C Ionization Suppression in Silicate Glasses Using SIMS, Goldschmidt (*Poster*).
- 2024 [15] Gleeson, M., Wieser, P., deVitre, C., **Shi, S.C.**, Millet, M.-A., Muir, D., Stock, M., Lissenberg, J., Persistent Magma Storage in the Mantle Across 2.5 Myrs of Ocean-Island Volcanism, GSA (*Talk*).
[14] **Shi, S.C.**, Wieser, P., Toth, N., Antoshechkina, P., Lehnert, K., mineralML: Leveraging Machine Learning for Probabilistic Mineral Classification, Gordon Research Seminar, Geochemistry of Mineral Deposits (*Invited Talk*).
[13] Tweedy, R., **Shi, S.C.**, Uno, K.T., Machine Learning Analysis of *n*-Alkanes from Woody and Grassy African Plants, NE Geobiology Conference (*Talk*).
- 2023 [12] **Shi, S.C.**, Wieser, P., Toth, N., Antoshechkina, P., Lehnert, K., MIN-ML: Leveraging Machine Learning for Probabilistic Mineral Classification in Geochemical Databases, AGU (*Talk*).
[11] Tweedy, R., **Shi, S.C.**, Uno, K.T., African Plant Functional Type Identification from *n*-Alkanes Chain Lengths via Non-Linear Methods, AGU (*Talk*).
[10] Bidgood, A., **Shi, S.C.**, Prabhu, A., Que, X., Twigg, H., Nulf, M., Using Supervised and Unsupervised Machine Learning Methods to Predict Missing Geochemical Data and Determine Geochemical Trends in Multielement Systems: Application to Sediment-Hosted Ore Deposits, AGU (*Poster*).
[9] Prabhu, A., Wong, M.L., Morrison, S.M.M., Ostroverkhova, A., Clark, M., Zhong, H., Prestgard, T.J., Li, W., Williams, J.R., **Shi, S.C.**, Mays, J., Hazen, R., From detecting agnostic biosignatures to characterizing chondrites: How network science is perfect for making scientific discoveries with geochemical data, AGU (*Invited Talk*).
[8] **Shi, S.C.**, Wieser, P., Toth, N., Antoshechkina, P., Lehnert, K., MIN-ML: A Machine Learning Framework for Exploring Mineral Relations and Classifying Common Igneous Minerals, Goldschmidt (*Invited Workshop Talk*).
[7] **Shi, S.C.**, Wieser, P., Lehnert, K., Profeta, L., MIN-ML: A Machine Learning Framework for Exploring Mineral Relations and Classifying Common Igneous Minerals, EGU (*Talk*).
- 2022 [6] Tweedy, R., **Shi, S.C.**, Uno, K.T., Grass in the Past: Eastern African Chemotaxonomy from Plant Wax *n*-alkanes, AGU (*Poster*).
[5] **Shi, S.C.**, Barth, A.C., Plank, T.A., Towbin, W.H., Flores, O., Arias, C.P., Magma stalling weakens eruption: Uncertainty quantification in thermometry and volatile measurements, VMSSG (*Talk*).
[4] Toth, N., **Shi, S.C.**, MacLennan, J., Automated petrography using machine learning, VMSSG (*Poster*).
- 2021 [3] **Shi, S.C.**, Barth, A.C., Plank, T.A., Towbin, W.H., Magma stalling weakens eruption, AGU (*Talk and ePoster*).
- 2018 [2] **Shi, S.C.**, Cerling, T.E., Uno, K.T., What plant is that? Chemotaxonomy from *n*-alkane molecular distributions of East African plants with implications for paleoecology, AGU (*Poster*).
[1] **Shi, S.C.**, Cerling, T.E., Uno, K.T., Resolving taxonomy with *n*-alkane molecular distributions of East African plants, Columbia University Chandler Society Research Symposium (*Invited Talk*).

LEADERSHIP AND SERVICE

- GORDON RESEARCH CONFERENCE ON THE GEOCHEMISTRY OF MINERAL DEPOSITS
2024-2026 **Gordon Research Seminar, Seminar Co-Chair**
· Elected co-chair for 2026 Gordon Research Seminar (early career); develop seminar program and select speakers
UNIVERSITY OF CAMBRIDGE, DEPARTMENT OF EARTH SCIENCE
2021-2022 **EDI, LGBTQ⁺, Geoscience in Context Working Groups, Committee Member**
· Perform outreach regarding equity, diversity, and inclusion
COLUMBIA UNIVERSITY
2018-2020 **Department of Earth Science Undergraduate Student Committee, Founder and Co-Chair**
· Developed seminar series to highlight climate and hazard research; initiated undergraduate involvement in equity outreach; spearheaded initiative to confer outstanding teaching awards to professors; led field geology trips alongside faculty
2016-2019 **Peer Health Exchange, Co-Coordinator and Diversity, Equity, and Inclusion Coordinator**
· Taught classes regarding physical and mental health/wellness to NYC high schoolers; recruited the most diverse chapter of 150 educators in organization history; developed and presented equity trainings; managed \$50k budget

TEACHING

	UNIVERSITY OF CALIFORNIA, BERKELEY
2025	Graduate Student Instructor for Genesis and Interpretation of Rocks
	COLUMBIA UNIVERSITY
2023-2024	Invited Speaker at Goldschmidt Workshop (Open Data in Geochemistry: Navigating Present Data Infrastructure); German Mineralogical Society Data Science in Geochemistry and Cosmochemistry · Develop Python module on applying supervised/unsupervised machine learning for classification of large mineral datasets Guest Lecturer for Earth’s Environmental Systems: Solid Earth · Present lecture on geochemical data and geoinformatics, develop Jupyter Notebook for visualizing and understanding petrologic trends in global mid-ocean ridge basalts
	UNIVERSITY OF CAMBRIDGE
2021-2022	Practical Demonstrator for Earth Sciences B (Second Year Module) · Lead demonstrations covering mineralogy, petrology, volcanology, and isotopes for second year undergraduates
2021	Field Demonstrator for Cornwall Field Geology Trip (6 Days)

CONFERENCES, DATATHONS, WORKSHOPS, FIELD WORK

2025	Stratigraphy and Earth History – Death Valley; Geologic Field Studies – Southwest U.S.
2024	AVERT Project Field Workshop – Poás Volcano, Costa Rica
2023	Mineral Informatics Datathon – Earth and Planets Laboratory, Carnegie Institute for Science Iceland Fieldwork – Cambridge Volcanology Group; Kenya Fieldwork – Uno Laboratory Goldschmidt Conference Workshop – Open Data in Geochemistry: Navigating Present Data Infrastructure Mineral Informatics Datathon – University of Idaho
2020	Goldschmidt Conference Workshop – Diffusion Chronometry

ADDITIONAL INFORMATION

Languages	English (native), Chinese (native), French (working proficiency)
Computation	Python (MC ³ , pandas, PyTorch, scikit-learn, TensorFlow), MATLAB, R, GitHub
Interests	Cycling, data sonification, music (production, journalism), fermentation, rowing