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

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)**

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, Cambridge — Pozzi Fund**
· Examined science and art intersections, resulting in data sonification of magma crystallization thermodynamic data

2020 **Columbia University**
· Phi Beta Kappa; Departmental Honors, awarded for academic and research excellence

Research

Data Science Fellow in Geoinformatics, Lamont-Doherty Earth Observatory

2022-2024 **Data Science Fellow Research**
· Develop 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

University of Cambridge

2021-2022 **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 (∼tens-hundreds of hours to ∼5 hours for each sample)
· Devise novel, open-source Bayesian MCMC algorithm to determine mineral chemistries with uncertainties by accounting for covariance in analytical and conversion factor uncertainties
· Model plagioclase diffusion profiles with finite-elements and Bayesian nested sampling to extract magma mixing timescales; evaluate spatiotemporal variability in plagioclase mineral chemistry across Iceland

Lamont-Doherty Earth Observatory, Columbia Climate School

2019-2021 **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 (CO_2, H_2O) 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

2024 [in preparation] **Shi, S.C.**, Wieser, P., Toth, N., Antoshechkina, P., Lehnert, K., mineralML: Leveraging Machine Learning for Probabilistic Mineral Classification in Geochemical Databases. *JGR: Machine Learning and Computation.*

2024 [in preparation] Toth, N., **Shi, S.C.**, Maclennan, J., Tung, P.Y., EDS Analysis for Petrology: A Probabilistic Framework. *JGR: Machine Learning and Computation.*

2024 [in preparation] 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.*

2024 [in press] 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 (H_2O, CO_2, S, Cl, F) Analyses and the C Ionization Efficiency Suppressing Effect of Water in Silicate Glasses. *Geostandards and Geoanalytical Research.*

2024 [in press] **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

2024 [16] **Shi, S.C.**, Antoshechkina, P., Lehnert, K., Profeta, L., Figueroa, J.D., Cao, S., Class, C., Wieser, P., Toth, N., Harnessing Flexible Search Tools and Machine Learning for Data-Driven Discovery, Goldschmidt 2024 (*Talk*).
[15] **Shi, S.C.**, Towbin, W.H., Plank, T.A., Barth, A.C., Rasmussen, D., Moussallam, Y., Lee, H., Menke, W., Quantifying H_2O and CO_2 Concentrations and Uncertainties with PyIRoGlass: An Open-Source Bayesian MCMC Algorithm for Fitting Baselines to Basaltic-Andesitic FTIR Spectra, Goldschmidt 2024 (*Talk*).
[14] **Shi, S.C.**, Wieser, P., Toth, N., Antoshechkina, P., Lehnert, K., mineralML: Leveraging Machine Learning for Probabilistic Mineral Classification, Gordon Research Seminar 2024, 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 GSA 2024 (*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 2023 (*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 2023 (*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 2023 (*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 2023 (*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 2023 (*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 2023 (*Talk*).

2022 [6] Tweedy, R., **Shi, S.C.**, Uno, K.T., Grass in the Past: Eastern African Chemotaxonomy from Plant Wax *n*-alkanes, AGU 2022 (*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, VMSG 2022 (*Talk*).
[4] Toth, N., **Shi, S.C.**, Maclennan, J., Automated petrography using machine learning, VMSG 2022 (*Poster*).

2021 [3] **Shi, S.C.**, Barth, A.C., Plank, T.A., Towbin, W.H., Magma stalling weakens eruption, AGU 2021 (*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 2018 (*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

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 school students; recruited applicants and actively developed leadership pipelines, achieving the most diverse chapter of 150 educators in history; developed and presented equity trainings to ensure cultural awareness and engagement; managed $50k budget

Teaching

Columbia University

2023-204 **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 to examine and classify 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)**

Datathons, Workshops, Field Work

2024 Poás Volcano, AVERT Project Field Workshop
2023 Earth and Planets Laboratory at Carnegie Institute for Science, Mineral Informatics Datathon
Iceland Fieldwork, Cambridge Volcanology Group; Kenya Fieldwork, Uno Laboratory
Goldschmidt Conference Workshop, Open Data in Geochemistry: Navigating Present Data Infrastructure
University of Idaho, Mineral Informatics Datathon
2020 Goldschmidt Conference, Diffusion Chronometry Workshop

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