Esther XU FEI

estherxufei@gmail.com 443-714-9400 estherxufei Estherrrrrxu

Skills

Programming: Python, SQL, R, MatLab, version control (Git), SLURM, LaTeX, shell script (bash)
Statistical: Time series analysis (ARIMA, ICA), dimension reduction (PCA), manifold learning (Isomap, diffusion map, UMAP, t-SNE), stochastic processes (MCMC, Gaussian Process), state space models (particle filters, particle MCMC)
Software: ParFlow, Gdal/ArcGIS/QGIS, Google Earth
Language: Chinese (native), English (fluent), and Spanish (intermediate)

Education

Ph.D. in Physical Hydrology | Johns Hopkins University (JHU) | Aug. 2024

Baltimore, MD | GPA: 3.92 / 4.00

- **M.Eng.** in Environmental Management and Economics, Department of Environmental Health and Engineering
- **M.Eng.** in Statistics and Statistical Learning, Department of Applied Mathematics and Statistics

M.S. in Hydrology | New Mexico Tech (NMT) | Aug. 2018

Socorro, NM | GPA: 3.93 / 4.00

- **Thesis** Estimation of Focused Recharge for New Mexico Using a Soil-Water-Balance Model: PyRANA
- Minor in Operational Research and Statistics

B.Eng. in (Petroleum) Resources Exploration Engineering | Yangtze University | Aug. 2017

Wuhan, Hubei, China | Dual degree program with NMT

B.S. in Earth Sciences with Geology option | NMT | Aug. 2016

Socorro, NM | GPA: 3.91 / 4.00

- Minor in Mathematics

Working Experiences

Applied Scientist III, Amazon.com, Inc.

Seattle, WA | Feb. 2024 – Present

- Leverage machine learning techniques to mitigate estimation biases in causal inference processes.
- Conduct impact assessments on millions of Amazon customers exposed to package marketing messages through semi-randomized experiments.

Applied Scientist Intern, Amazon.com, Inc.

Seattle, WA | May 2023 – Aug. 2023

- Identified 5 distinct personas from a high-dimensional dataset of over 400k customers and 2k features using PySpark.
- Integrated continuous business metrics into an automated and interpretable personalization framework, overseeing the coordination of various functional components, including dimensionality reduction (UFS), high-dimensional data transformation, and latent archetype space construction (WoE, VAE), clustering (IHC-KNN).
- Developed a prompt in-context learning system to enhance the interpretability of the personalization pipeline.
- Leveraged AWS SageMaker and EC2 to deploy large language models, including Falcon-40B and Llama-2-7B.
- Collaborated with a diverse team of Product Managers, scientists, Software Development Engineers, and Data Engineers from various geographical regions.
- Currently preparing publication for submission.

Research Assistant, Johns Hopkins University (JHU)

Baltimore, MD | Aug. 2018 – Present

- Conduct research in the area of Hydrology and Applied Mathematics focusing on modeling hydrological time series and time series classification through literature research, developing models, analyzing data, and academic research communication
- Investigate flow path and nutrient cycling in natural watersheds by using mathematical methods to model complicated time series
- Develop skills in Python programming, time series modeling, isotope analyzer operation, and field experiments

Research Projects

Automatic air quality monitoring at Curtis Bay

Baltimore, MD | Oct. 2023 - Present

- Spearhead a team consisting of 3 students and 2 faculty members in the creation of a real-time air quality monitoring system, employing machine learning and computer vision technologies
- Engineer picture labeling mechanisms utilizing optical character recognition (OCR) and image processing pipelines to streamline air quality monitoring procedures.
- Design and implement a computer vision model capable of detecting and categorizing coal mine activity, complemented by a machine learning model to correlate air quality fluctuations with coal mining operations, facilitating support for community air quality regulations.

SigROCKET: a scalable time series classification method

Baltimore, MD | Mar. 2023 - Present

- Developed a machine learning model capable of classifying time series data characterized by gaps in time, large fluctuations in magnitude, and similar ranges of fluctuations. Real-world applications include distinguishing between different types of diseases based on ICU measurements.
- Integrated the non-linear feature extraction method, Signature, with Multi-Rocket to construct a scalable time series classification algorithm.
- Devised a scalable algorithm suitable for analyzing long multivariate time series data, achieving state-of-the-art classification performance with up to a 35% increase in AUC, while reducing computational costs. This was accomplished by combining signature transformation with Multi-Rocket.
- One publication currently under preparation.

Uncertainty estimation of transit time distribution in a bi-modal hydrologic response watershed

Baltimore, MD | Aug. 2021 - Present

- Disaggregated weekly bulk samples into 6-hourly intervals using Gaussian Process regression. Additionally, propagated a 28x downscaled input and its 95% uncertainty bound using Gaussian Process regression, passing hypothesis tests at a 0.99 confidence level. Finally, propagated the downscaled input through a fluid transport model.
- Developed a model to characterize the transit time distribution of a watershed, enabling the understanding of how this watershed transports, stores, and releases substances.

Bayesian uncertainty quantification on MESAS model

Baltimore, MD | Sep. 2018 - Present

- Enhanced stream water solute concentration estimation by implementing a data-driven approach using local linear piecewise StorAge Selection (SAS) functions, eliminating the need for a-priori assumptions.
- Strengthened a numerical fluid transport model with a multiscale adaptive kernel algorithm, facilitating non-parametric estimation of hidden states at flexible local scales through various non-linear statistical inference methods. Developed a bash pipeline for conducting experiments on a high-performance computing grid powered by SLURM. Submitted software publication demonstrating a 10x reduction in cumulative numerical error for a 4-year test dataset.
- Devised a particle Markov Chain Monte Carlo (MCMC) framework for propagating input and output uncertainty,

enabling informed risk control by providing accurate inference about the structure of complex dynamic systems.
- Conducted simultaneous uncertainty quantification on multiple sources, including input/output time series and black-box model structure.
- Authored one software publication, with one publication currently under preparation.

Construct coarse representation of subsurface soil-rock interface

Baltimore, MD | May 2020 - Feb.2024
- Derive effective coarse-scale representation of permeability at subsurface permeability contrast to facilitate demand for detailed data (requires intensive drilling) in previous fill-and-spill modeling.
- Test proof-derived anisotropic permeability tensor from realizations generated from virtual truth based on the Richards equation using the ParFlow model.
- Developed bash pipeline on high performance computing grid MARCC to process parallel computing tasks for more than 30,000 computational hours.

Theoretical optimal benchmarking in time series classification

Baltimore, MD | May 2022 - Feb. 2023
- Establish theoretical optimal benchmark to evaluate SOTA time series classification (TSC) methods (random forest, ROCKET, neural network) for stochastic process
- Provide synthetic dataset (Ornstein-Uhlenbeck processes, different potentials, Brownian motion + constant drift, Opinion dynamics) for systematical testing on TSC methods
- Authored in one preprint publication.

Novel end-member identification model, CHEMMA

Baltimore, MD | May 2018 - May 2021
- Advanced traditional end-member identification method by building unsupervised data-driven model in Python, Convex-Hull End-Member Mixing Analysis (CHEMMA) by reducing data annotation 100%
- Successfully identified 3 field-measured end-members combining ConvexHull-NMF and constrained Kmeans; Reduce streamwater chemistry variation 6x on each end-members; Achieve same result using 50% less data
- Published one first author publication on Hydrology and Earth System Sciences; Model is recognized as SOTA on invited review paper
- Code publicly available on GitHub

Data analysis on hydrological connectivity of Bonneville salt flats

Baltimore, MD | Oct. 2018 - Dec. 2018
- Identified surface deformation through Principal Component Analysis (PCA) and Independent Component Analysis (ICA) on remotely sensed dataset (InSAR) of Bonneville salt flat

Statewide groundwater recharge estimation (M.S. Thesis)

Socorro, NM | Jun. 2016 - Aug. 2018
- Developed Python-programmed groundwater recharge model to estimate rate and distribution of groundwater recharge for entire state of New Mexico with team of 9; Improved algorithm in giving estimation at karst landscape with error less than 10%; Cooperated with New Mexico Tech evapotranspiration (ET) research group to improve estimation of ET and total available water in root zone
- Estimated precipitation-runoff relationship by building linear regression model with threshold, and reduced overland flow by 7 times.
- Processed 2T GIS files using Gdal, ArcGIS, and QGIS; Acquired soil physical property data from USDA NRCS soil database STATSGO and SSURGO
- Authored one publication summarizing project as first author (under preparation) and one masters thesis

Awards and Honors

M. Gordon Wolman Fellowship | Aug. 2023 - May 2024

Doctoral Leadership Award, JHU EHE | May 2023
Natalie M. Lorenz Anderson Fellowship | Aug. 2021 - May 2022
CUAHSI student travel grant | Jan. 2019 - Oct. 2019
Edwin D. and Rachel Lowthian Endowed Fellowship | Aug. 2018 - Aug. 2019
UC Berkeley workshop scholarship | May 2019
Environmental Health and Engineering Student Organization Travel Grant | Apr. 2019
Lee and Albert H. Halff Doctoral Student Award | Aug. 2018
New Mexico Tech Graduate Student Study Travel Grant | Sep. 2017
New Mexico Tech Honor Roll | Aug. 2014 - May 2016
Durtche Geophysics Award (Best geophysics student of the year) | May 2016
NMGS student Fall Field Conference Scholarship | Oct. 2015
Carlsbad Mineral and Gem Society Award (Best geology student of the year) | May 2015
Best debater in Yangtze University Debate Finals (1/20) | Mar. 2014
Yangtze University Scholarship | Sep. 2012 - Jun. 2014

Publications

[1] Ciaran Harman and **Esther Xu Fei**. mesas.py v1.0: a flexible python package for modeling solute transport and transit times using storage selection functions. *Geoscientific Model Development*, 17(2):477–495, 2024. doi: 10.5194/gmd-17-477-2024. URL https://gmd.copernicus.org/articles/17/477/2024/.

[2] Zehong Zhang, Fei Lu, **Esther Xu Fei**, Terry Lyons, Yannis Kevrekidis, and Tom Woolf. Benchmarking optimality of time series classification methods in distinguishing diffusions. *arXiv preprint arXiv:2301.13112*, 2023.

[3] Jianyu Fu, Bingjun Liu, Weiguang Wang, and **Esther Xu Fei**. Evaluating main drivers of runoff changes across china from 1956 to 2000 by using different budyko-based elasticity methods. *Journal of Environmental Management*, 329:117070, 2023.

[4] **Esther Xu Fei** and Ciaran Harman. A data-driven method for estimating the composition of end-members from stream water chemistry time series. *Hydrology and Earth System Sciences*, 26(8):1977–1991, 2022.

[5] Jianyu Fu, Weiguang Wang, Benjamin Zaitchik, Wanshu Nie, **Esther Xu Fei**, Scot M Miller, and Ciaran J Harman. Critical role of irrigation efficiency for cropland expansion in western china arid agroecosystems. *Earth's Future*, 10(9):e2022EF002955, 2022.

[6] **Fei Xu**. *Estimation of Focused Recharge for New Mexico Using a Soil-Water-Balance Model: PyRANA*. PhD thesis, 2018. URL https://www.proquest.com/dissertations-theses/estimation-focused-recharge-new-mexico-using-soil/docview/2093745891/se-2.

Teaching and Mentoring Experiences

Teaching Assistant

Data Analytics in Environmental Health and Engineering | Jan. 2024 - May 2024
Data Analytics in Environmental Health and Engineering | Jan. 2023 - May 2023
Data Analytics in Environmental Health and Engineering | Jan. 2022 - May 2022

Data Analytics in Environmental Health and Engineering | Jan. 2021 - May 2021
Landscape Hydrology and Watershed Analysis | Jan. 2020 - May 2020
Landscape Hydrology and Watershed Analysis | Aug. 2018 - Dec. 2018
Introduction to Fluid Mechanics | Aug. 2018 - Dec. 2018

Mentor

Ph.D. Candidate from Cornell University, Nish Singh | Nov. 2023 - Present
M.S.E. student from Johns Hopkins University, Sakshi Labhane | Jun. 2021 - Dec. 2021
Undergraduate student, Kayla Ostrow | Dec. 2019 - Mar. 2020
High school student, Julia Alumbro | Aug. 2019 - Dec. 2019

Leadership and Services

Leadership

Primary Convenor | Mar. 2023 - Dec. 2023
Session H23D, AGU Fall Meeting 2024

Student representative | Jan. 2023 - Aug. 2024
Cross-Institutional Student Advisory Committee, JHU, Baltimore, MD

Treasurer | May. 2022 - May 2024
Graduate Representative Organization (GRO), JHU, Baltimore, MD

President | May 2022 - May 2023
EHE Student Service Organization (EHESO), JHU, Baltimore, MD

Student speaker | Aug. 2022
EHE Welcome Ceremony, JHU, Baltimore, MD

Local conference organizer | Jun. 2021 - May 2022
Astrobiology Graduate Conference (AbGradCon) 2022, Washington, D.C.

Secretary | Aug. 2021 - May 2022
GRO, JHU, Baltimore, MD

Student speaker | May. 2021
EHE Graduation Ceremony, JHU, Baltimore, MD

President-elect | May 2021 - May 2022
EHESO, JHU, Baltimore, MD

Lab Assistant | Aug. 2018 - May 2021
Landscape Hydrology Lab, JHU, Baltimore, MD

Ph.D. Representative | Aug. 2020 - May 2021
EHESO, JHU, Baltimore, MD

Co-host | Oct. 2016
NM Statewide Water Assessment Workshop, Socorro, NM

Resident Assistant | Aug. 2015 - Aug. 2016
Residential Life, New Mexico Tech, Socorro, NM

Chief Editor | Sep. 2013 - Aug. 2014
School Magazine, Yangtze University, Hubei, Wuhan, China

Volunteer

Hopper | Sep. 2023
Grace Hopper Celebration 2023, Orlando, FL

Coffee Hour Host | Dec. 2020
EHESO, JHU, Baltimore, MD

Science Tour Guide | Feb. 2015 - May 2015
Magdalena Ridge Observatory, Magdalena, NM

Presentations

Oral presentations

Guess lecturer for Data Analytics, JHU, Baltimore, MD | Feb. 2023

- Covered lectures for linear and non-linear regression, logistic regression, boostrap, cross validation, model selection and regularization

2022 American Geophysical Union Fall Meeting, Chicago, IL | Dec. 2022

- Dynamic uncertainty quantification of catchment transit time and StorAge Selection distributions using an adaptive non-parametric Bayesian framework

2022 Frontiers in Hydrology, San Juan, PR | Jun. 2022

- CHEMMA: a method for estimating end-member source composition from mixture data alone

2021 American Geophysical Union Fall Meeting, New Orleans, LA | Dec. 2021

- Estimating the transit time distribution in a forested watershed with a bimodal hydrologic response using Multi-scale Estimation of StorAge Selection function (MESAS)

JHU Environmental Health and Engineering department seminar, Baltimore, MD | Feb. 2021

- Where does the water in your cup come from?

JHU Environmental Health and Engineering department seminar, Baltimore, MD | Nov. 2019

- CHEMMA 101: Introduction to Convex Hull End Member Mixing Analysis

JHU Environmental Health and Engineering department seminar, Baltimore, MD | Sep. 2018

- Estimation of focused recharge for New Mexico using a soil-water-balance model: PyRANA

2017 American Geophysical Union Fall Meeting, New Orleans, LA | Dec. 2017

- Statewide groundwater recharge modeling in New Mexico

NM Statewide Water Assessment Workshop, Socorro, NM | Oct. 2016

- Water estimation matters

Poster presentations

2023 American Geophysical Union Fall Meeting | Dec. 2023

- SigROCKET: a scalable time series classification algorithm using path signature and random convolution kernel
- An adaptive Bayesian approach for stochastic dynamic system uncertainty quantification with applications to noisy, incomplete, or excessively-smoothed data

2020 American Geophysical Union Fall Meeting | Dec. 2020

- Can fill-and-spill subsurface flow be represented by a moisture-dependent anisotropic permeability tensor in Richards' equation-based models with coarse spatial resolution?

2019 American Geophysical Union Fall Meeting | Dec. 2019

- Learning from the data: manifold learning in interpreting tracers of the landscape hydrologic system

2018 American Geophysical Union Fall Meeting | Dec. 2018

- High-resolution statewide groundwater recharge estimation by soil water balance in New Mexico

62^{nd} New Mexico Water Conference Aug. 2017

- Efforts on calibration and validation of modeling groundwater recharge in New Mexico

61^{st} New Mexico Water Conference Oct. 2016

- Modeling focused recharge through ephemeral streams in New Mexico

Participated Workshops

Princeton GPU Hackathon 2022 | Jun. 2022
Princeton University, Princeton, NJ

HydroML | May 2022
Pennsylvania State University, University Park, PA

Advanced short course: integrated simulation of watershed systems using ParFlow | Oct. 2019
University of Arizona, Tucson, AZ

4th annual Communicating Science Conference in Chicago | Aug. 2019
Northwestern University, Evanston, IL

Short course: integrated simulation of watershed systems using ParFlow | May 2019
Colorado School of Mines, Golden, CO

Workshop on critical timescales of hydrologic transport | May 2019
University of California, Berkeley, Berkeley, CA

Short course: environmental models and Bayesian methods | Mar. 2019
University of Waterloo, Waterloo, Ontario, CA

Master class: advanced techniques in watershed science | Jan. 2019
Biosphere2, Oracle, AZ