Takuya Kurihana

tkurihana at fujitsu.com *change at to @
Knoxville, TN

Skills Python, Fortran 90/95, C, C++, C#, MPI, OpenMP, Shell Script, MySQL, Java, HTML, SciLab, Matlab, Go, Microsoft Office, Latex, Jira, Git, Notion.
ML/DL libraries Tensorflow, Keras, Pytorch, Horovod, Scikit-learn, Spark-mllib, Weight and Biases
Libraries AWS (EC2, DynamoDB, Lambda, S3, AutoScaling), GCP (Vertex AI), Spark, Kubernetes, Docker, Parsl, Pachyderm, Dask, Elasticsearch, HPC Scheduler (SLURM, Torque, Cobalt)

RESEARCH / TECHNICAL INTEREST

Self-supervised learning, Physics-informed machine learning, High-Performance Computing, Generative AI, Clustering, Cloud computing, Remote sensing, Earth System/Environmental science

EXPERIENCE

Oak Ridge National Laboratory — Oak Ridge, TN
Postdoctoral Research Assistant — May, 2024 – March, 2025

- **Developing generative model-based seasonal forecasting systems** to improve the predictability of seasonal and sub-seasonal weather forecasts, under the guidance of post-doctoral advisor Moetasim Ashfaq;
- **Fine-tuning weather foundation models** to enhance predictability and sub-seasonal forecasts, and developing Earth-observation models to optimize computational resources for multi-decadal ocean satellite data analysis. [EGU'24, IGARSS'24, ORBIT, IBM-NASA].
- **Working on scalable multi-facility science workflows** to automate data and labor-intensive Earth-observation ML pipelines into a federated and seamless process, orchestrated with workflow applications [SC'24 Xloop].

University of Chicago — Chicago, Illinois
Graduate Research Assistant — August, 2018 – March, 2024

- **Developed Rotationally Invariant Cloud Clustering [RICC; TGRS'21]**, a self-supervised clustering framework, to **create intermediate representations that are agnostic to rotations of images**, improving clustering performance of grouping similar object patterns and textures. RICC trains with 20TB MODIS satellite imagery and I **scale up the training/inference to 128 NVIDIA A100 GPUs [eScience'21] and >1024 CPU nodes.**
- Created the AI-driven Cloud Classification Atlas [**AICCA; MDPI'22 Editor's choice**], a novel classification dataset produced by applying RICC 23 years of MODIS Aqua and Terra instruments – 850 **Terabytes or 198 million ocean cloud imagery, to identify robust and meaningful 42 clusters of cloud patterns [NeurIPS-W'22]**. AICCA provides a compact form 56GB of labels and parameters to climate scientists, **democratizing access to the Big Science Data.**

IBM Research — Yorktown, New York
Research scientist — June – September, 2023

- Develop physics-informed AI for emulating high-resolution climate simulations

Design and train a physics-informed deep neural network by developing cutting-edge computer vision techniques (i.e., generative model and vision transformer) to upscale coarse-resolution climate data into high-resolution ones by $\times 9$ times in horizontal resolution. Specifically, I work on a super-resolution of three-dimensional wind simulation data to create a physics-informed emulator for accelerating simulation speeds and decreasing computational costs.

The preliminary model development and the results are summarized into an abstract and submitted to the American Geophysical Union Fall Meeting 2023, and NeurIPS 2023 Workshop: Tackling Climate Change with Machine Learning.

During the internship, I learn about business opportunities through meetings with clients as a staff research scientist. Also, I have collaborated with >6 and more staff research scientists to develop solutions for the super-resolution project and joined meetings about geospatial foundation models to discuss the downstream tasks and developments with other IBM scientists in Yorktown, Japan, Brazil, and Europe office.

Frontier Development Laboratory Mountain view, California
Research scientist June – August, 2022

- Build a fast and reliable contaminant plume prediction under future climate scenarios

Developed (1) an end-to-end deep Siamese autoencoder via online clustering assignments that optimizes reconstruction and clustering loss terms simultaneously to generate novel climate patterns from CMIP5 simulations, and (2) improve an U–Net integrated Fourier neural operator (U-FNO) by integrating physics-informed loss terms for groundwater transport and flow problems, to achieve multi-scale digital twin application. I worked with domain scientists to design two different deep learning algorithms, and employed data ingestion and training pipeline with tensorflow and horovod to scale them on Google Cloud Platform.

Woven Planet Holdings Tokyo, Japan
Software engineer July – October, 2021

- Conduct benchmarks and optimization of the high-definition map pipeline

Developed a benchmark tool running in a Kubernetes with Prometheus that measured the throughput from the high-definition map (HD map) creation pipelines. Conducted extensive amount of experiments to investigate bottleneck of pipelines, optimizing the configuration of production workflow using ~200TB of satellite imagery.

- Develop dynamic autoscaling operator

Developed and deployed an autoscaler as a Kubernetes operator that adjusts the number of k8s pods based on performance of pipelines. This autoscaler addressed suboptimal pipeline performances especially when multiple pipelines are competing the cloud resources. There is no conflict with the cloud native autoscale policy, e.g., AWS AutoScaler, as the scope of proposed autoscaler dynamically allocate optimal number of k8s pods of different pipelines running concurrently on cluster computers. The operator increased the efficiency of resource allocation by at most 500%.

Japan Meteorological Research Institute Tsukuba, Japan
Research Assistant June, 2016–July, 2018

- Study on uncertainty of cumulonimbus initiation and development using particle filter

Developed parallelized Python analytical tools for dual-polarization doppler radar observation for statistical tests and assimilated simulation results.

Researched the observation error correlation in doppler radar products and simulations; Presented the results at Japan Meteorological Society 2018.

- Study on the optimal perturbation method for ensemble data assimilation

  Researched impacts of off-diagonal elements in transform matrix under the ensemble transformation of ensemble Kalman filter via conducting observing system simulation experiments (OSSE) on SPEEDY-LETKF for 50 ensemble members. I contributed to present the results for several peer-reviewed conferences and published into a book.

  Developed Fortran95/bash-script tools for spectrum analysis, perturbation technique, and ensemble analysis for OSSE results.

Mitsubishi UFJ Morgan Stanley Securities Tokyo, Japan
Financial Engineer February–March, 2017

Researched weak-/strong-scaling experiments on 10 000 cores to optimize performance of a bond option pricing model

Developed an computation-time efficient algorithm for derivative price computation, and reduced the computation time by 50%.

Mizuho Securities Tokyo, Japan
Quantitative Analyst August, 2017

Researched case studies of volume weighted average price (VWAP) for derivative option pricing.

RIKEN Advanced Institute for Computational Science Kobe, Japan
Graduate Research Assistant July–August, 2017

Developed a super-observation (aggregation of dense observation to the nearest grid point of numerical simulation) algorithm to accelerate aggregation of NASA Global Precipitation Measurement product for the Nonhydrostatic ICosahedral Atmospheric Model (NICAM) i.e., find a closest grid location against a satellite observation among O(100 000) and more number of grid points. The adaption of computing efficient superobservation approach allows NICAM to simulate more realistic precipitation estimation.

Rice University Houston, TX
Research Assistant August–September, 2016

Developed Kalman filter and expectation-maximization algorithm for a diffusion-advection nonlinear partial differential equation

EDUCATION

University of Chicago Chicago, IL
PhD in Computer Science GPA 3.868 March 2024

- Research area: Deep learning (Unsupervised-/Semi-supervised learning application in climate science), Machine learning, Clustering, High performance computing. Cloud dynamics.

  (1) Developed the rotation-invariant optimization function to improve feature representation in autoencoder. The rotation-invariant feature improves clustering of object patterns and textures regardless of their orientation of inputs without any assumptions concerning artificial categories, addressing misclassification depended on their rotations of object.

  (2) Developed an unsupervised cloud clustering framework to train on 20TB dataset processed from MODIS satellite products, discovering 42 intermediate cloud categories underexplored in conventional cloud classifications, and scaled the training up to 128 NVIDIA A100 GPUs and $>$1024 CPU nodes.

(3) Created the AI-driven Cloud Classification Atlas (**AICCA**), a novel classification dataset produced by applying rotation-invariant autoencoder and hierarchical clustering to identify robust and meaningful clusters of cloud patterns based on 23 years of MODIS Aqua and Terra instruments – 850 Terabytes or 198 million ocean cloud imagery.

These research outputs were summarized into four papers accepted in Climate Informatics 2019, IEEE Transactions on Geoscience and Remote Sensing, MDPI Remote Sensing, and NeurIPS Workshop. These results were also presented in AGU Fall Meeting 2019, 2022, and 2023.

- Honors: Heiwa Nakajima Global Scholarship, MDPI editor's choice (AICCA: AI-Driven Cloud Classification Atlas)

MS in Computer Science GPA 3.60 December 2020

- Specializations: (1) Distributed tensorflow training, (2) Neuromorphic computing, (3) Auto-scaling strategies, (4) Deep learning on IoT device

(1) Benchmark task for distributed tensorflow over three cluster computers. Evaluate impact of TCP /IP network configuration (jitter, latency and packet loss) on the distributed tensorflow training.

(2) Coded a spiking based unsupervised neural network on neuromorphic computing simulator. Use a hand-digits dataset (MNIST) to analyze different spiking signals on their bottleneck layer via PCA, revealing that different digit-class shows different spiking patterns.

(3) Developed automatic scaling strategies aiming at efficient resource managements for a Python parallel scripting library (Parsl). The strategies improved the resource utilization by 30% by the reallocation of resources at unnecessary workers.

(4) Modified tensorflow, Bazel and OS applications for ARM ODROID XU4 to deploy trained deep learning models. Code a Python wrapper for C/C++ energy consumption library, and evaluate consumed energies by different kernels.

University of Tsukuba Tsukuba, Japan
MS in Meteorology GPA 3.92 March 2021

- Specializations: Data assimilation, Numerical optimization, Numerical weather prediction.

Developed a satellite-observation pre-processing algorithm that reduced the computation time by 98% than that of greedy algorithm against dense satellite observations.

Evaluated the impact of multiple pre-processing algorithms to assimilated results, showing that the proposed outperformed the error statistics in analysis.

University of Tsukuba Tsukuba, Japan
BS in Meteorology GPA 3.60 March 2017

- Rice University, Research Internship Program, August–September, 2016
- University of California Berkeley, Summer program, June–August, 2015
- Honors: Nakatani - RIES Foundation Fellowship, Best Presentation Awards in International Student Science Forum Ho Chi Minh City, Second Presentation Prize in Urayasu City English Presentation Competition

ACTIVITIES

Teaching Assistant Chicago, IL

- Teaching assistant for Inclusive technologies January–March, 2022

Mentor 15 undergraduate students to advise their weekly class-project proposals and inclusive technologies applications. One of my mentored team won the best award in class.

- Teaching assistant for Computer Science with Applications 2 January–March, 2021

  Manage two laboratory classes up to 40 undergraduate students to instruct python (padas, matplotlib), web-scraping, regular expression, bashscript, database and algorithms. Hold office hours and question dashboard to assist students about class materials.

- Teaching assistant for Cloud computing April–June, 2020

  Instruct Amazon web service (EC2, Lambda, EBS, DynamoDB) and open office hours to assist up to 60 master students.

Internship mentor Chicago, IL
Summer student supervisor

- Student mentor in METCALF program June–September, 2020

  Supervised one undergraduate student to build an image annotation tool based on CVAT library for MODIS calibrated radiance product..

- Student mentor in Center for Data and Computing program June–September, 2019

  Supervised one master and two undergraduate students to develop automation libraries of MODIS satellite produces' pre-processing for unsupervised cloud clustering.
  Supervised an undergraduate student to compare clustering analysis (HAC, K Means++, Spectral, and DBSCAN) for outputs from our developed cloud classification framework.

SELECTED PUBLICATION

1. **Kurihana, T.**, Ashfaq, M., Horan, M., & Aslam, Z., Knowledge-guided multimodal deep neural network toward seasonal precipitation forecasting in Central Southwest Asia. *AGU Fall Meeting 2024* Abstracts 2024, GC23D-0296.
2. **Kurihana, T.**, Skluzacek, T. J., da Silva, R. F., & Anantharaj, V. (2024). Scalable multi-facility workflows for artificial intelligence applications in climate research. *The International Conference for High-Performance Computing, Networking, Storage, and Analysis, Proceedings of the 6th Annual Workshop on Extreme-Scale Experiment-in-the-Loop Computing (XLOOP 2024)*. https://doi.org/10.1109/SCW63240.2024.00266.
3. Padovani, G., Anantharaj, V., Sacco, L., **Kurihana, T.**, Bunino, M., Tsolaki, K., Girone, M., Antonio, F., Sopranzetti, C., Fronza, M., & Fiore, S. (2024). A software ecosystem for multi-level provenance management in large-scale scientific workflows for AI applications. *The International Conference for High-Performance Computing, Networking, Storage, and Analysis, Proceedings of the 19th Workshop on Workflows in Support of Large-Scale Science (WORKS 2024)*. https://doi.org/10.1109/SCW63240.2024.00253.
4. Anantharaj, V., **Kurihana, T.**, Dash, S., Padovani, G., & Fiore, S. (2024). Exploring vision transformers on the Frontier supercomputer for remote sensing and geoscientific applications. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2024)*. https://doi.org/10.1109/IGARSS53475.2024.10640929.
5. **Kurihana, T.**, Mastilovic, I., Wang, L., Meray, A., Praveen, S., Xu, Z., Memarzadeh, M., Lavin, A., & Wainwright, H. (2024). Identifying climate patterns using clustering autoencoder techniques. *Artificial Intelligence for the Earth Systems*. https://doi.org/10.1175/AIES-D-23-0035.1.
6. Meray, A., Wang, L., **Kurihana, T.**, Mastilovic, I., Praveen, S., Xu, Z., Memarzadeh, M., Lavin, A., & Wainwright, H. (2023). Physics-informed surrogate modeling for supporting climate resilience at groundwater contamination sites. *Computers & Geosciences*, 183, 105508. https://doi.org/10.1016/j.cageo.2023.105508.

7. **Kurihana, T.**, Yeo, K., Szwarcman, D., Elmegreen, B., Mukkavilli, K., Schmude, J., & Klein, L. (2023). A 3D super-resolution of wind fields via physics-informed pixel-wise self-attention generative adversarial network. *The Thirty-Seven Annual Conference on Neural Information Processing Systems (NeurIPS): Tackling Climate Change with Machine Learning*. arXiv preprint arXiv:2312.13212.
8. **Kurihana, T.**, Franke, J., Foster, I., Wang, Z., & Moyer, E. (2022). Insight into cloud processes from unsupervised classification with a rotationally invariant autoencoder. *The Thirty-six Annual Conference on Neural Information Processing Systems (NeurIPS): Machine Learning and the Physical Sciences*. arXiv preprint arXiv:2211.00860.
9. Wang, L., **Kurihana, T.**, Meray, A., Mastilovic, I., Praveen, S., Xu, Z., Memarzadeh, M., Lavin, A., & Wainwright, H. (2022). Multi-scale Digital Twin: Developing a fast and physics-informed surrogate model for groundwater contamination with uncertain climate models. *The Thirty-six Annual Conference on Neural Information Processing Systems (NeurIPS): Machine Learning and the Physical Sciences*. arXiv preprint arXiv:2211.10884.
10. **Kurihana, T.**, Moyer, E. J., & Foster, I. (2022). AICCA: AI-driven cloud classification atlas. *Remote Sensing*, 14(22), 5690. https://doi.org/10.3390/rs14225690.
11. **Kurihana, T.**, Moyer, E., Willett, R., Gilton, D., & Foster, I. (2021). Data-driven cloud clustering via a rotationally invariant autoencoder. *IEEE Transactions on Geoscience and Remote Sensing*, 1-25. https://doi.org/10.1109/TGRS.2021.3098008.
12. **Kurihana, T.**, Moyer, E., Willett, R., Gilton, D., & Foster, I. (2021). Cloud Clustering Over January 2003 via Scalable Rotationally Invariant Autoencoder. *IEEE 17th International Conference on eScience (eScience)*, 253-254.
13. **Kurihana, T.**, Foster, I., Willett, R., Jenkins, S., Koenig, K., Werman, R., Lourenco, R. B., Neo, C., & Moyer, E. (2019). Cloud characterization with unsupervised deep learning. *Proceedings of the Climate Informatics Workshop 2019*, Paris. https://arxiv.org/abs/2209.15585.
14. **Kurihana, T.**, Foster, I., Moyer, E. J., Willett, R., Maire, M., Jenkins, S., Koenig, K., and Werman, R. Cloud Characterization With Deep Learning II. *AGU Fall Meeting 2019* Abstracts 2019, A53H-03.