

# Albert M. W. Yau

---

[LinkedIn](#)

[GitHub](#)

[Google Scholar](#)

## Education

<b>School of Marine and Atmospheric Sciences, Stony Brook University</b>	MPhil, <i>Atmospheric Science</i>	2020
Dissertation: Finding Storm Track Activity Metrics That Are Highly Correlated with Weather Impacts		
Advisor: Prof. Edmund K.M. Chang		
<b>Scipps Institution of Oceanography, UC San Diego</b>	MS, <i>Oceanography</i>	2009
Advisor: Prof. Myrl C. Hendershott		
<b>The Chinese University of Hong Kong</b>	BSc, <i>Theoretical Physics</i>	2008
Minors: <i>Computer Science, Mathematics</i>		
<b>University of Toronto</b>	Exchange	2006
<b>UC Berkeley</b>	Exchange	2005

## Fellowship and Scholarships

<b>California Institute of Technology</b>	Summer Undergraduate Research Exchange (SURE) Fellowship	2007
<b>University of Toronto</b>	Yasumoto International Exchange Scholarship	2006
<b>The Chinese University of Hong Kong</b>	CN Yang Scholarship	2005

## Work Experience

**R1 RCM | Senior Software Engineer** 2023–Present

- Architected and implemented Spark Structured Streaming pipelines to replicate in-memory datastores to Databricks, processing 3M+ records/minute with < 60s p99 latency, reducing annual compute costs by \$3.5M.
- Resolved concurrency bottlenecks of an AI charge categorization service by optimizing Spark DAGs to increase performance by 15x, and reconfiguring Spark clusters to reduce latency from 5 minutes to < 30 seconds.
- Designed and built a serverless data lake API, migrating web applications to query Databricks Delta Lake tables directly; eliminated a Redis caching layer based on load testing to simplify infrastructure.

**dv01 (Acquired by Fitch Group) | Data Engineer (Infrastructure)** 2018–2023

- Owned the end-to-end functional Scala/Spark data infrastructure for this Series A startup; architected reusable ingestion patterns to accelerate customer-facing integrations.
- Led a PB-scale storage migration from Azure to Google Cloud Platform; modernized legacy architecture to cloud-native services: SQL Server to BigQuery, Airflow to Cloud Composer, Spark to Dataproc, and VMs to Kubernetes.
- Executed a zero-downtime migration of a monolithic Scala codebase from Spark 2 to 3 ahead of Dataproc end-of-life, resolving cross-environment dependency conflicts across Python, Scala and R services.
- Prototyped from a Hack Day project, productionized a text parsing pipeline using custom Airflow operators and FastAPI microservices to extract unstructured PDF data; scaled capacity to 200+ securitizations.
- Designed and implemented the disaster recovery plan for financial data via multi-region replication and tiered backups to ensure high availability and data durability.

## Research Experience

**Stony Brook University** | Graduate Student Researcher 2012–2018

Advisor: Prof. Edmund K.M. Chang, School of Marine and Atmospheric Sciences

- Provisioned and administered bare-metal Linux storage clusters (300+ TB capacity) using hardware RAIDs for high-throughput on-premises climate model analysis.
- Built data pipelines to aggregate 200 years of global climate model simulations across 30+ international institutions, standardizing datasets with different formats and spatial resolutions into a unified data store.
- Developed EOF-CCA statistical frameworks and storm track activity metrics to predict extreme weather patterns; publications were cited in the United Nations IPCC Sixth Assessment Report (AR6).

**National Center for Atmospheric Research** | Data Science Intern 2015

Advisors: Dr. Kevin Paul and John Dennis, Application Scalability and Performance Group

- Implemented a high-performance cyclone tracking program (PyStormTracker) within 10 weeks during the Summer Internships in Parallel Computational Science (SIParCS).
- Designed accelerated cyclone detecting algorithms using NumPy and SciPy; scaled the tracking algorithm across 4,096 cores on the Yellowstone supercomputer using MPI4Py.

**Coastal Environments / Scripps Institution of Oceanography** | Oceanographer 2010–2011

Advisors: Dr. Hany Elwany and Prof. Myrl C. Hendershott, Integrative Oceanography Division

- Modeled coastal processes, including beach erosion, tidal dynamics, and far-field tsunamis; validated lagoon hydrodynamic models against collected GPS drifter data.
- Isolated low-amplitude, short-period sea-level fluctuation modes in the Santa Barbara Channel by applying spectral and complex-EOF analysis to noisy bottom pressure sensor data.

**Academia Sinica** | Visiting Researcher 2008

Advisor: Dr. Danie M.C. Liang, Research Center for Environmental Changes

- Processed NASA GISS global climate model outputs.

**California Institute of Technology** | Summer Undergraduate Research Exchange (SURE) Fellow 2007

Advisor: Prof. Yuk L. Yung, Division of Geological and Planetary Sciences

- Analyzed solar cycle response in IPCC global climate models; results presented at the AGU Fall Meeting.

**The Chinese University of Hong Kong** | HPC Support Assistant / Student Researcher 2005–2008

- Optimized C/C++ and Fortran scientific workloads by tuning compiler flags and OpenMP/MPI parallelization.
- Deployed HPC clusters over InfiniBand and GbE; constructed a campus-wide computational grid (CUGrid) utilizing Globus Toolkit and HDFS; integrated with the Pacific Rim and Grid Middleware Assembly (PRAGMA).
- Configured, initialized and ran Weather Research and Forecasting (WRF) model simulations on HPC clusters at CUHK and HKO to model gravity waves downstream from Lantau Peak over Hong Kong International Airport.

**Hong Kong Observatory** | Research Intern 2006

Advisor: Dr. Ping-Wah Li, Aviation Weather Services

- Ingested LIDAR data from Hong Kong International Airport into the Local Analysis and Prediction System (LAPS).

# Publications and Presentations

## Journal Publications

Citations: ~300 as of March 2026 / h-index: 5 (Google Scholar)

1. Chang, E.K.M., **A.M.W. Yau**, and R. Zhang, 2022: Finding Storm Track Activity Metrics That Are Highly Correlated with Weather Impacts. Part II: Estimating Precipitation Change Associated with Projected Storm Track Change over Europe. *Journal of Climate*, **35**, 2423–2440, <https://doi.org/10.1175/JCLI-D-21-0259.1>.
2. **Yau, A.M.W.**, and E.K.M. Chang, 2020: Finding Storm Track Activity Metrics That Are Highly Correlated with Weather Impacts. Part I: Frameworks for Evaluation and Accumulated Track Activity. *Journal of Climate*, **33**, 10169–10186, <https://doi.org/10.1175/JCLI-D-20-0393.1>.
3. Chang, E.K.M., and **A.M.W. Yau**, 2016: Northern Hemisphere winter storm track trends since 1959 derived from multiple reanalysis datasets. *Climate Dynamics*, **47**, 1435–1454, <https://doi.org/10.1007/s00382-015-2911-8>.
4. Chang, E.K.M., C.G. Ma, C. Zheng, and **A.M.W. Yau**, 2016: Observed and projected decrease in Northern Hemisphere extratropical cyclone activity in summer and its impacts on maximum temperature. *Geophysical Research Letters*, **43**, 2200–2208, <https://doi.org/10.1002/2016GL068172>.
5. Chang, E.K.M., C. Zheng, P. Lanigan, **A.M.W. Yau**, and J.D. Neelin, 2015: Significant modulation of variability and projected change in California winter precipitation by extratropical cyclone activity. *Geophysical Research Letters*, **42**, 5983–5991. <https://doi.org/10.1002/2015GL064424>.

\* 3. and 4. cited in the United Nations IPCC Sixth Assessment Report (AR6).

## In Preparation

6. **Yau, A.M.W.**, and M.C. Hendershott: Short Period Sea-level Fluctuations in the Santa Barbara Channel.  
\* Currently validating spectral and Complex-EOF analysis results using a shallow-water wave model.

## Selected Poster Presentations

- **Yau, A.M.W.**, K. Paul, and J. Dennis, 2016: PyStormTracker: A parallel object-oriented cyclone tracker in Python. *96<sup>th</sup> American Meteorological Society Meeting*, New Orleans, LA, 454.
- **Yau, A.M.W.**, Y. Guo, and E.K.M. Chang, 2013: Impacts of background field removal on the seasonal cycle and trend of cyclone statistics. *AGU Fall Meeting 2013*, San Francisco, CA, A31B-0051.
- Tung, K., **M. Yau**, K. Li, R. Shia, F. Li, D. Waliser, and Y. Yung, 2007: Solar-cycle response in global climate models assessed by IPCC AR4. *AGU Fall Meeting 2007*, San Francisco, CA, GC43A-0935.

## Workshops

- 20<sup>th</sup> Annual Community Earth System Model (CESM) Workshop, NCAR, 2015, Breckenridge, CO.
- Community Atmosphere Model (CAM) Tutorial, NCAR, 2009, Boulder, CO.
- Pacific Rim and Grid Middleware Assembly (PRAGMA):
  - PRAGMA 18, 2010, San Diego, CA.
  - PRAGMA 14, 2008, Taichung, Taiwan.

## Technical Skills

**Languages:** Python, Scala, C#, C/C++, Fortran, MATLAB, Bash **Cloud:** Google Cloud Platform, Azure, Databricks  
**Data:** SQL, Spark, Delta Lake, BigQuery **Frameworks:** NumPy, SciPy, Numba, Dask, Xarray, PyTorch, MPI, OpenMP  
**DevOps:** Kubernetes, Docker, Terraform, Airflow **Systems:** Linux (Debian/Ubuntu/WSL), Proxmox, ZFS

## Projects

### PyStormTracker | PyPI

2015–2018; 2026–Present

- Originally developed for distributed HPC environments at NCAR using MPI4Py, with recent integrations of Dask multi-processing to parallelize high-resolution climate data workloads.
- Refactored core algorithms using vectorized NumPy arrays and Numba JIT-compiled kernels, with an 11x speed-up on serial workloads. The bottleneck is now in the centralized linking algorithm, which is being investigated.
- Upgraded to Python 3.11+ with modern typing and linting, complete unit and integration test suite, including legacy regression tests, and multi-arch Docker and PyPI publishing pipelines using GitHub Actions.
- Actively developing improved tracking algorithms, spectral preprocessing and storm activity metrics.

### Parallel FFT and SHT | SBU CSE613 Parallel Programming Project

2013

- Integrated Intel Cilk++ parallel SDK with FFTW and ccSHT libraries for benchmarks.
- Achieved an 80% parallel efficiency for large  $512^3$  3D Fast Fourier Transforms, outperforming the 50% baseline of the original multi-threaded FFTW implementation.
- Achieved a 32% peak parallel efficiency for Spherical Harmonics Transforms on 1,024 zonal wave numbers.

### XDDD.org | Founder

2002–2005

- Built a web hosting and webmail service, with a physical server co-located in the HKNet data center.
- Created a secure multi-tenant Linux/Apache/MySQL/PHP (LAMP) stack on Red Hat Linux, with strict user isolation via chroot, suPHP, and disk quotas.

### Homelab

2019–Present

- Built a dual WAN 10GbE fiber network and Proxmox bare-metal virtualization environment with ZFS storage.
- Managed IoT devices and containerized workloads on Raspberry Pis and NVIDIA Jetson Nano.
- Contributed to open-source IoT projects: IoTaWatt (home electricity monitoring), graphs1090 (flight tracking).

## Interests

- Astrophotography
- Deep Learning
- Home automation
- Software-defined radio (SDR)