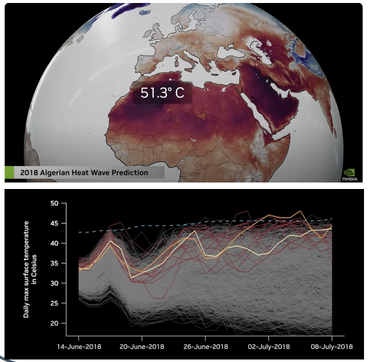
CAlibrated and Systematic Characterization, Attribution, and Detection of Extremes (CASCADE) SFA (**PI**: Bill Collins – LBNL, **TCM**: Travis O'Brien – IU)

To advance understanding of singular climate extremes known as low-likelihood, high-impact events (LLHIs), the drivers that cause them, and the evolution of these drivers in warmer climates

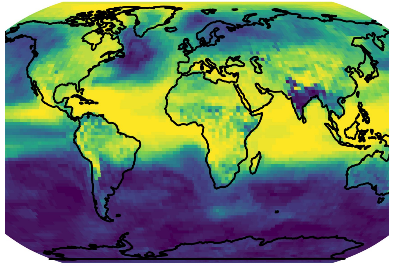
Ensembles for LLHIs



NVIDIA:

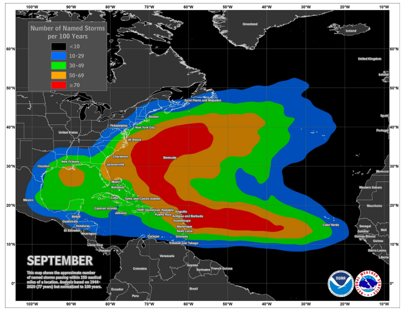
How can ML-based emulation help increase fidelity of simulated LLHIs?

LLHIs in the Observational Record



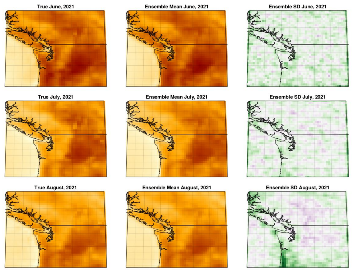
How can the observational record be leveraged to improve statistical and physical understanding of LLHIs?

Prediction and Projection of LLHIs



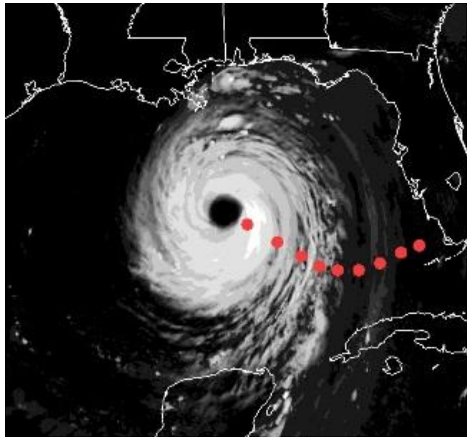
What are the sources of LLHI predictability that can provide early warning of LLHIs at S2I & S2D timescales?

Machine Learning & Computational Infrastructure

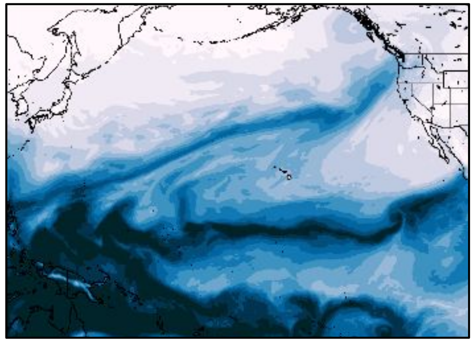


What machine learning and computational tools are needed to advance understanding of LLHIs?

Key research accomplishments during CASCADE3

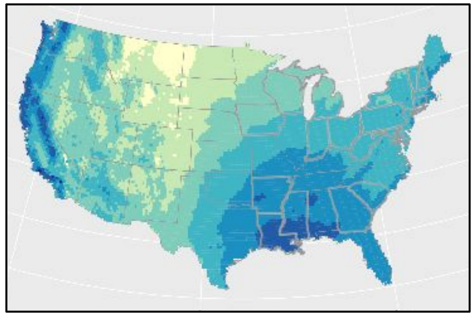


Extremes in the Observational Record
- Method and dataset development: tools for analysis of extremes at impact-relevant scales
- Characterization, detection, and attribution of extremes using in situ observations
- Evaluation of extremes in climate model simulations



Subseasonal to Multidecadal Variability in Extremes
- Forced responses in extremes versus responses to natural variability
- Multiscale interactions between extremes and their precursors
- Physical mechanisms driving variability and change in extremes

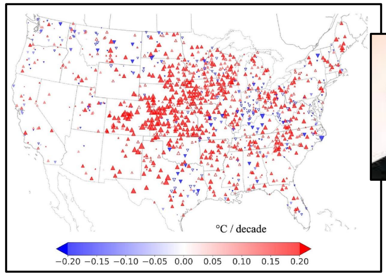
Observing and Modeling Extremes at Their Native Scales
- Multiscale processes and feedbacks required for hydroclimate extremes in climate simulations
- Toward theory-based understanding of multiscale extremes
- Multiscale convective processes and teleconnections



Key Highlights from FY22/23 during CASCADE3

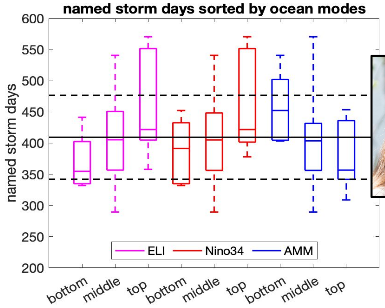
Key Findings:

- Trends of 0.3°C/decade occur in volatility of extreme daily temps. for Central and Eastern United States.



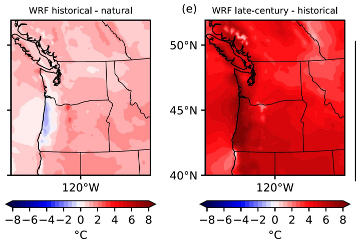
Weather and Climate Extremes

- La Niña / El Niño is associated with reduced / increased global TC frequency.



Geophysical Research Letters

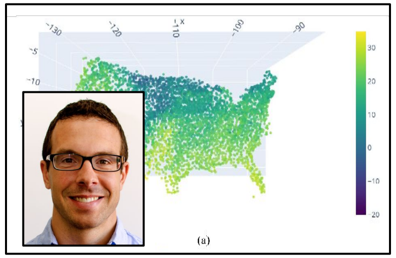
- Further warming could lead to a ~5°C increase in heatwaves by the end of 21st C.



Geophysical Research Letters

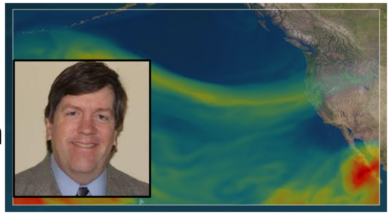
Broader Implications:

- We develop infinitely-scalable, probabilistic ML methods for analyzing huge data volumes on newest DOE HPC.



Nature Scientific Reports

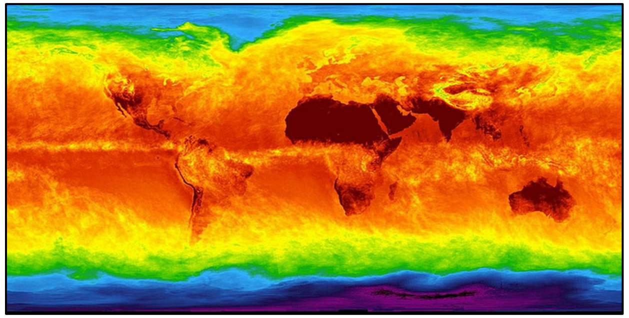
- Combining topology with ML yields an alternative and efficient storm detection algorithm useful in very large climate model datasets.



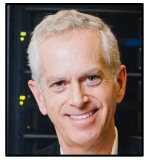
Big Earth Data Analytics

Future Directions:

- FourCastNet emulator for Low-likelihood High Impact Extremes
- Huge Ensembles (HENS) of climate projections



FourCastNet, Pathak et al. (2022), 0.25°, ~1,000,000 Pixels, ViT+AFNO



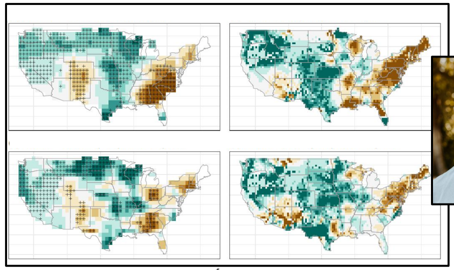

CASCADE Team Highlights

- **19 papers** (3 in Nature/Science families)
- **1 book chapters**
- **65 presentations** (11 invited)
- **Science Plan for CASCADE4 developed.**

Key Highlights from FY23/24 during CASCADE3

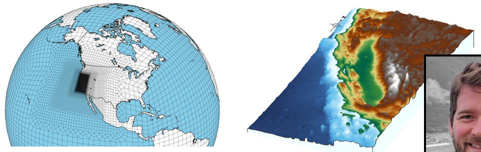
Key Findings:

- Decreases in aerosols mask GHG-forced change in precipitation



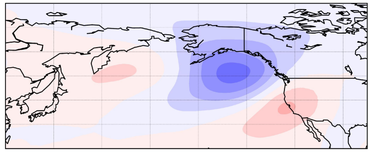
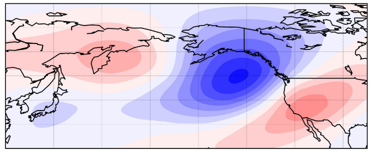
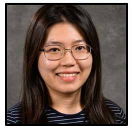

nature

- Regional refinement yields high quality simulation of the 1997 New Year's flood in E3SM




Journal of Advances in Modeling Earth Systems

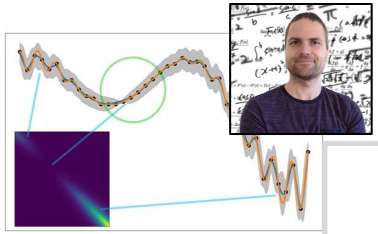
- Back-to-back ARs will likely become more common as GHGs increase

nature portfolio **communications** earth & environment

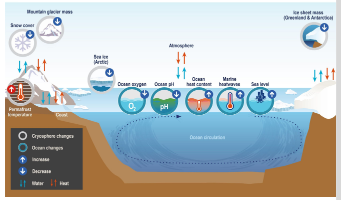
Broader Implications:

- We develop new statistics-backed machine learning approaches for nonstationary, noisy data



nature portfolio SCIENTIFIC REPORTS

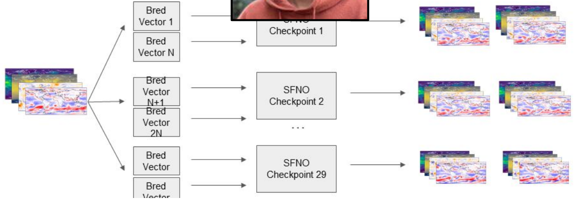
- Contributions to the National Climate Assessment enable climate-informed decision making



National Climate Assessment 2023

Future Directions:

- Forthcoming papers document near-miss extremes in the un-precedented 2 PB HENS ensemble
- Continue mutually beneficial collaborations with NERSC and NVIDIA

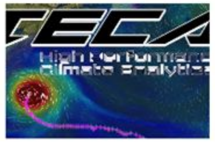
HENS ensemble workflow for the NVIDIA FourCastNet model enabled by tight collaboration with NERSC and NVIDIA, as well as two major ALCC allocations

CASCADE Team Highlights

- **34 papers** (7 in Nature/Science families)
- **2 book chapters**
- **93 presentations** (47 invited)
- **Science Plan Reviewed** (Dec 2023) **and Accepted** (April 2024)

Key technical accomplishments up thru CASCADE3

see https://cascade.lbl.gov/software-products



Toolkit for Extreme Climate Analysis (TECA)

TECA is a collection of climate analysis algorithms geared toward extreme event detection and tracking implemented in a scalable parallel framework. The code has been successfully deployed and run at massive scales on current DOE supercomputers. TECA's core is written in modern C++ and exploits MPI + X parallelism where X is one of threads, OpenMP, or GPUs. The framework supports a number of parallel design patterns including distributed data parallelism and map-reduce. While modern C++ delivers the highest performance, Python bindings make the code approachable and easy to use.

Loring, Burlen; Dufek, Amanda; Elbashandy, Abdelrahman; Johnson, Jeffrey N; Keen, Noel; Krishnan, Harinarayan; O'Brien, Travis A; Prabhat; and the CASCADE SFA. Copyright 2015: The Regents of the University of California, through Lawrence Berkeley National Laboratory

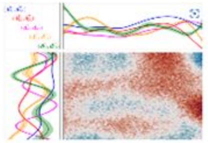
Get package at GitHub →

gp2Scale: Large Scale Exact Gaussian Processes

The Gaussian Process (GP) is arguably the leading uncertainty-quantification tool at the disposal of machine learning engineers and spatial statisticians. Its one major disadvantage is scalability which limits competitiveness with other machine learning methodologies. The Python library gp2Scale -- part of the fvgp package (pip install fvgp) -- makes exact GPs scalable to millions of data points via kernels that let the algorithm discover naturally occurring sparsity in datasets.

Nature, https://www.nature.com/articles/s41598-023-30062-8 Marcus M. Noack, Harinarayan Krishnan, Mark D. Risser & Kristofer G. Reyes. https://doi.org/10.1038/s41598-023-30062-8

Available on Github →

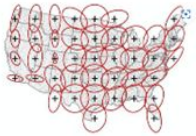


The BayesianSVD package for Julia

A Bayesian mixed-effects statistical model where the random effect is parameterized by a singular value decomposition (SVD). The matrices of the SVD representation are estimated using a Bayesian framework, enabling uncertainty quantification of the singular values and vectors in the decomposition.

North, J. S., Risser, M. D., Breidt, F. J. (2023) "A flexible class of priors for conducting posterior inference on structured orthonormal matrices". arXiv. https://doi.org/10.48550/arXiv.2307.13627

Link: view on GitHub →



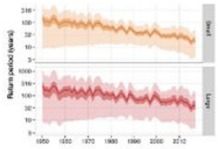
The convoSPAT package for R

Fits convolution-based nonstationary Gaussian process models to point-referenced spatial data via local likelihood techniques. Also provided are functions to fit stationary spatial models for comparison, calculate the kriging predictor and standard errors, and create various plots to visualize nonstationarity.

Risser, M.D., Calder, C.A. (2017). Local Likelihood Estimation for Covariance Functions with Spatially-Varying Parameters: The convoSPAT Package for R. Journal of Statistical Software, 81(14), 1-32. DOI: 10.18637/jss.v081.i14

Available on GitHub at https://github.com/cran/convoSPAT

Learn more at CRAN →

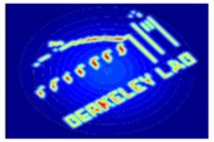


The climextRemes package for R and Python

climextRemes is an R and Python package for extreme value analysis of climate data, including generalized extreme value and peaks-over-threshold (using the point process approach) models, as well as tools for estimating risk ratios with uncertainty. Latest revision - 0.3.0: August 2022.

Paciorek C, Stone D, Wehner M (2018). "Quantifying statistical uncertainty in the attribution of human influence on severe weather." Weather and Climate Extremes, 20:69-80. doi: 10.1016/j.wace.2018.01.002

Bitbucket climextRemes →

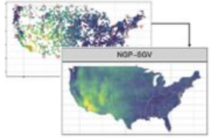


fastKDE

Fast, objective kernel density estimation cited in over 140 papers and used in over 35 software projects across the sciences.

O'Brien, T. A., Kashinath, K., Cavanaugh, N. R., Collins, W. D. & O'Brien, J. P. A fast and objective multidimensional kernel density estimation method: fastKDE. Comput. Stat. Data Anal. 101, 148-160 (2016). http://dx.doi.org/10.1016/j.csda.2016.02.014

Available at GitHub →



The BayesNSGP package for R

Enables off-the-shelf functionality for fully Bayesian, nonstationary Gaussian process modeling using a personal laptop. We furthermore implement approximate Gaussian process inference to account for moderately large spatial data sets. Bayesian inference and posterior prediction is carried out using Markov chain Monte Carlo methods.

Risser, M. D., & Turek, D. (2020). Bayesian inference for high-dimensional nonstationary Gaussian processes. arXiv preprint arXiv:1910.14101.

Learn more at CRAN →

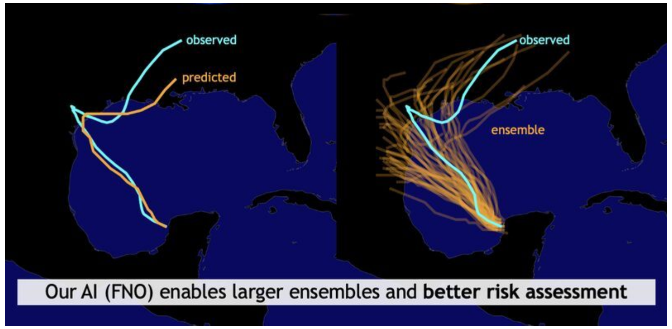
Impact of CASCADE Capabilities & Tools:

- Continued development of community scientific software products
- Cited in 150+ publications
- International impacts across disciplines; used in:
  - statistics, astronomy, AI, automation, engineering, particle physics, etc.
  - USA, China, England, France, Russia, Australia, Germany, etc.
- An increasingly engaged community: pull requests, issue/bug reports, feature requests, etc.
- Leadership in community dataset production & curation:
  - C20C+: 400+ publications enabled
  - Atmospheric River Tracking Method Intercomparison Project (collab w/ Catalyst)
    - 22 publications enabled
  - HighResMIP Data Lake @ NERSC
    - 7 publications enabled

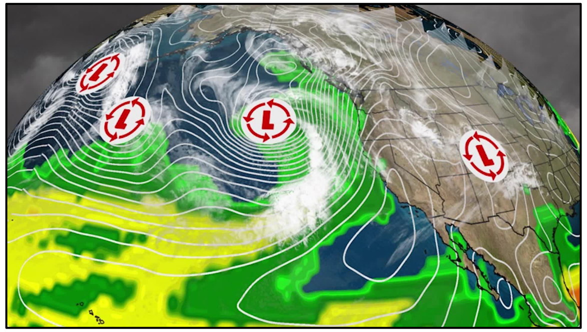
Key Scientific Objectives (3-5 years) for CASCADE4

Ensemble Driver Framework for Low-Likelihood High-Impact (LLHI) Extremes

- Framework to switch focus from LLHIs to their drivers
- Study of LLHIs in HighResMIP from CMIP6 using TECA
- Characterization of LLHIs in E3SM
- Study of LLHIs in Huge Ensembles built with FourCastNet
- Tests of FourCastNet using slow-moving tropical cyclones



Anima Anandkumar, NVIDIA and Caltech

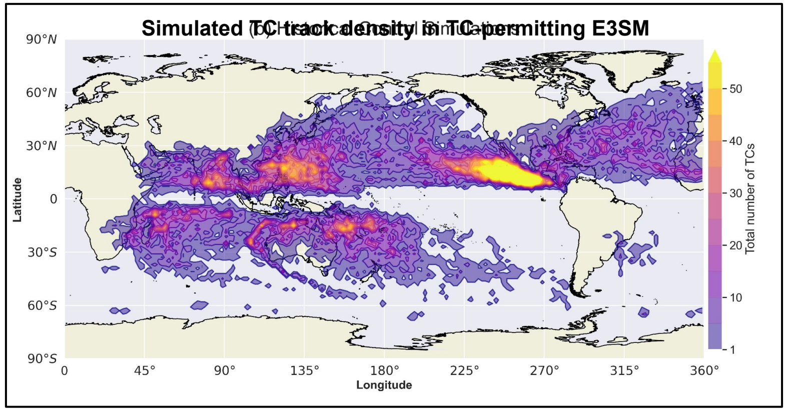


LLHIs in the Observational Record

- Observed LLHI probabilities, uncertainties, and drivers
- Statistical and AI/ML methods for D&A of observed changes in LLHIs
- Empirical weather ensembles for interpreting observed LLHIs

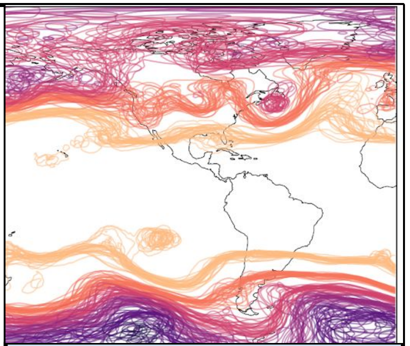
Prediction and Projection of LLHIs

- Limits of LLHI predictability: E3SM extreme event reforecasts
- Variability and predictability of LLHIs in the Western US
- Sources of predictability in LLHI TCs and TC seasons
- Future change in severe storms in convection-permitting dynamical climate models and machine learning models



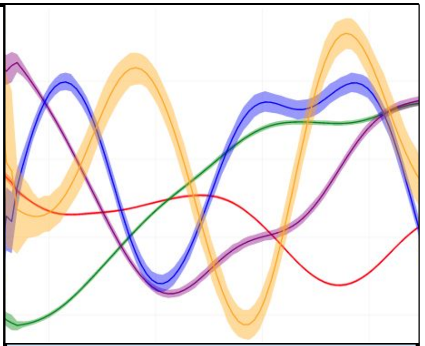
Key Technical Objectives (3-5 years) for CASCADE4

- Drive NVidia FourCastNet with TECA: detecting ARs, TCs, blocking systems, etc.
- Advance and improve FourCastNet, collaborate w/ NVidia
- Explore huge ensembles of hindcast simulations



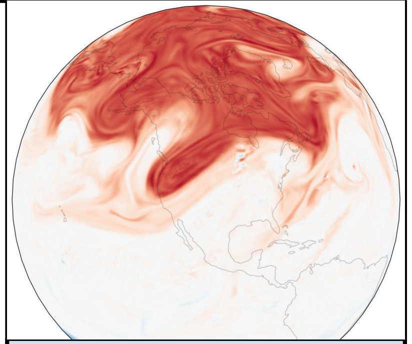
Huge Ensembles of NN Simulations

- Explore Bayesian PCA and SVD methods
- Develop ensemble gridded datasets of temperature and precipitation based on BPCA



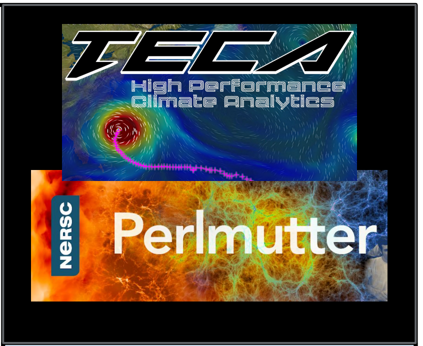
Feature Extraction & Dim. Reduction w/ UQ

- Implement new stochastic foundation models based on Gaussian processes
- Develop new covariance formulations based on idealized physics (e.g., quasigeostrophy)



Stochastic Foundation Models for Earth Science

- Enhance GPU support in all TECA components
- Facilitate ML tasks w/ TECA: from training to inference
- Incorporate ML layers in TECA pipelines (e.g., FourCastNet)



TECA for ML and GPUs

Partners for CASCADE4

NVIDIA

Existing collaboration:
- Huge Ensemble of Neural-network Simulations (HENS)

Near-term collaborations:
- New HENS with future versions of Spherical FourCastNet (SFNO)
- Extension of HENS to mesoscale precipitation

NCAR catalyst

Existing & near-term collaborations:
- Extreme event hindcast ensembles w/ E3SM
- ARTMIP

PCMDI – Earth System Model Evaluation Project

Existing & near-term collaborations:
- Metrics for evaluating ML climate emulators
- AR metrics in PMP
- Coordinated Model Evaluation Capabilities (CMEC)

NERSC

Existing & near-term collaborations:
- Toolkit for Extreme Climate Analysis (TECA) development
- ML climate model development & evaluation

HyperFACETS – U.S. Department of Energy | Office of Science – A Framework for Improving Analysis and Modeling of Earth System and Intersectoral Dynamics at Regional Scales

Existing & near-term collaborations:
- Mountain hydroclimate & extremes
- Decision-relevant extremes metrics

Conferences led by CASCADE4

Machine Learning for Actionable Climate Science

Exploiting Machine Learning to Enhance Earth System Modeling and Analysis Across Scales

Gordon Research Conference

June 22 - 27, 2025 • Bryant University

Exploiting Machine Learning to Enhance Earth System Modeling and Analysis Across Scales

This GRC conference will explore how to best push the frontiers of ML beyond state-of-the-art approaches, especially in

1. developing hybrid Earth system models with greater fidelity,
2. providing capabilities for climate extremes through large ensembles with emulators as well as enhancing detection and attribution methods, and
3. advancing climate model analysis and benchmarking.

This interdisciplinary conference brings together ML and climate scientists, as well as the private sector, to accelerate progress towards actionable climate science.

Organizers

Veronika Eyring
Chair

William Collins
Co-Chair

Pierre Gentine
Vice Chair

Laure Zanna
Co-Vice Chair

Chris Lofholm
Conference Operations Associate

CASCADE Team Presentations at this Meeting







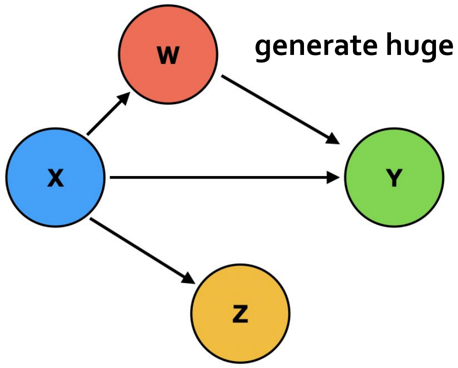


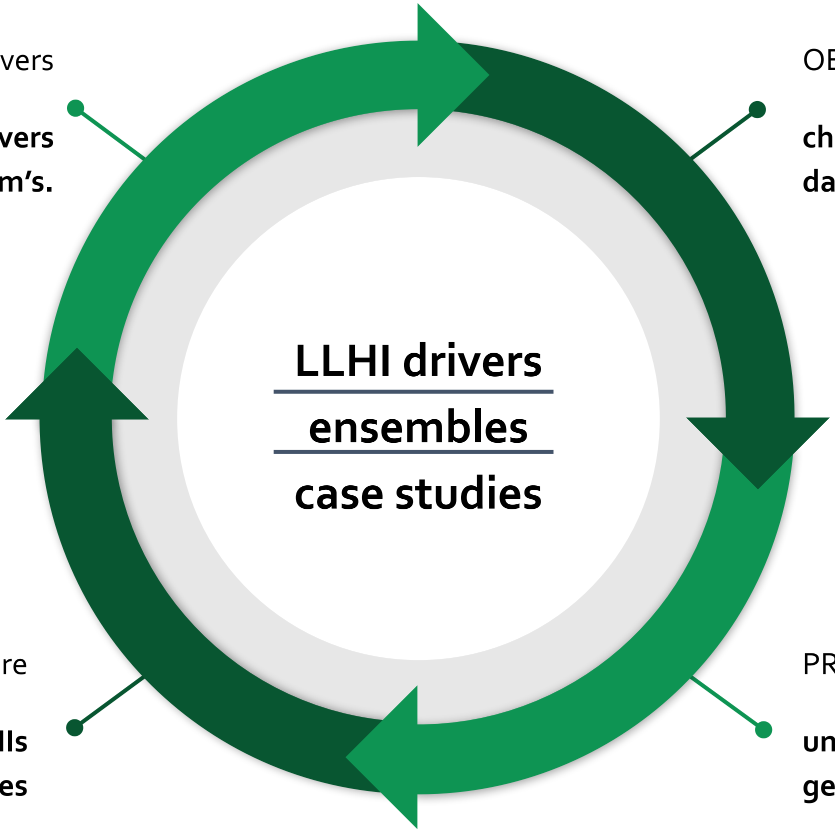

Lead Presenter	Topic	Session	Day
Emily Bercos-Hickey	**Tropical Cyclone Tornadoes**	Poster Session G	Wednesday (075)
Bill Collins	**Huge Ensembles for LLHIs**	Oral Session 2	Thursday
Derrick Danso	Effects on SLR and TCs on surges		
Ankur Mahesh	**Huge Ensembles using SFNO**	Plenary Session V	Friday
Josh North	**Variability UQ on LLHIs**	Poster Session G	Wednesday (074)
Travis O'Brien	**US ARs assoc. w/Baroclinic Waves**	Poster Session G	Thursday (061)
Chris Patricola	**Coastal Impacts of Tropical Cyclones**	Plenary Session V	Friday
Alan Rhoades	**LLHI Snowmelt Events**	Poster Session A	Thursday (006)
Mark Risser	**LLHI Historical Heatwaves**	Poster Session G	Wednesday (068)
Yang Zhou	**AR River Clusters**	Poster Session G	Wednesday (064)

CASCADE4 Summary:
Advance LLHI simulation, understanding, attribution and prediction

ENS | Ensembles for LLHI Drivers

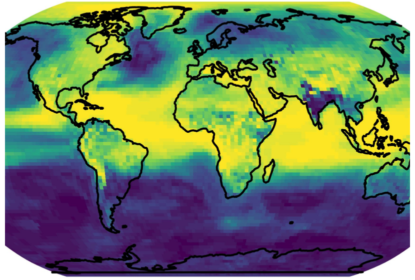
diagnose LLHI drivers
generate huge ensembles of NN sim's.

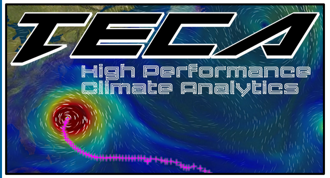




OBS | LLHIs in Observations

characterize LLHI driver statistics
data-driven emulation of LLHIs





INF | ML and Comp. Infrastructure

develop ML tools for LLHIs
detect drivers in ensembles

PRE | LLHI Predictability

understand physics of LLHI drivers
generate E3SM reforecasts of LLHIs