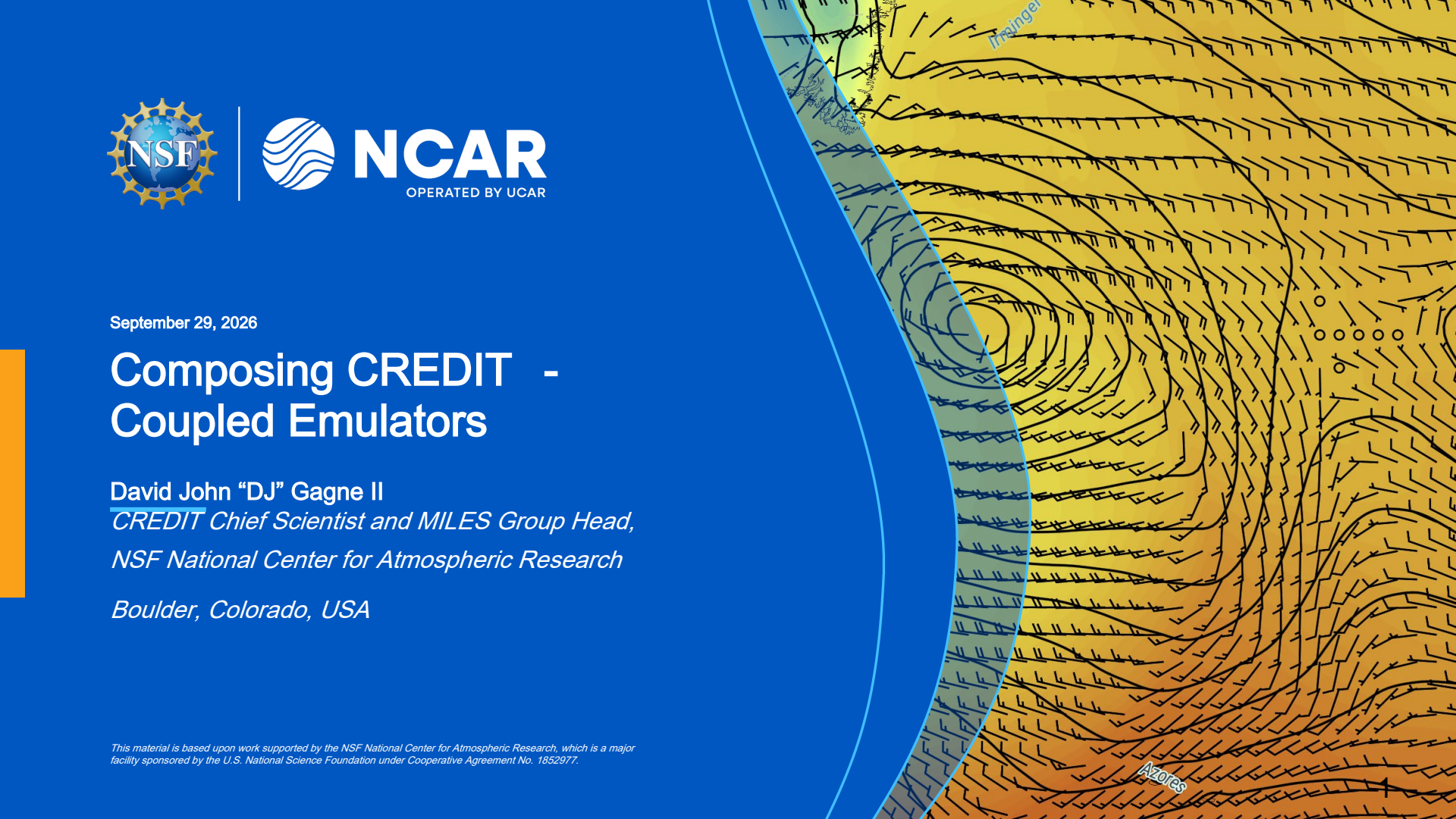


September 29, 2026

Composing CREDIT - Coupled Emulators

David John "DJ" Gagne II
CREDIT Chief Scientist and MILES Group Head,
NSF National Center for Atmospheric Research
Boulder, Colorado, USA

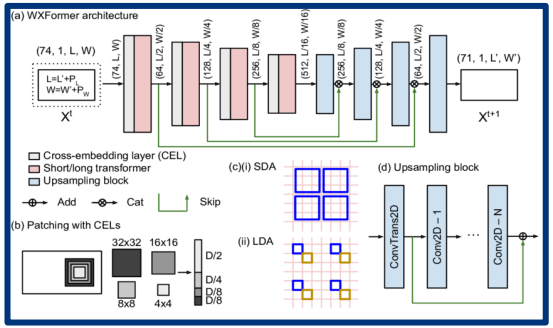
This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

Motivation: Unpacking CREDIT's Development



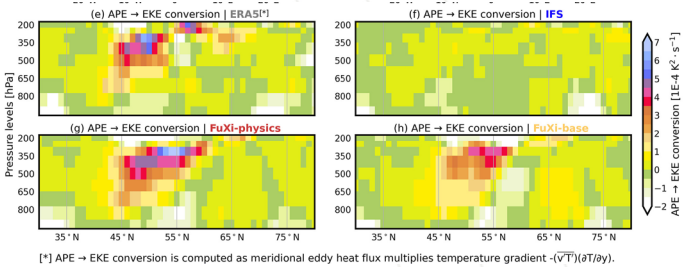
Summer - Fall 2023

Derecho finally came online. Wanted to train our own AI NWP emulators. Model - level ERA5 on GLADE. DJ, John, Will map out CREDIT plans over Domino's pizza.



Winter - Fall 2024

Will builds initial CREDIT data pipeline. John experiments with different mishmashes of architectures, builds WXFormer. DJ validates models and debugs physics issues. Initial WXFormer release.



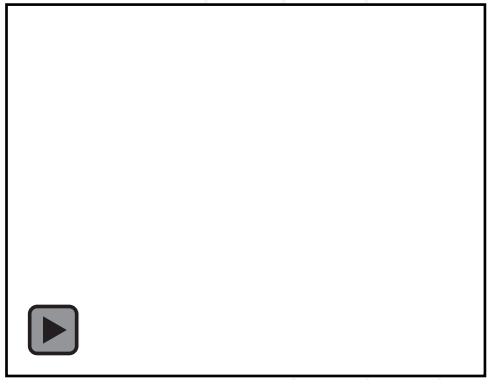
2025

Kyle Sha finds improvements with physics constraints. Will develops CAMulator . John builds stochastic decomposition layers for ensembles.

Running Up CREDIT Technical Debt

- CREDIT team generated lots of exciting research results really fast…
- But the code became more brittle and fragmented.
  - Component duplication
  - Missing metadata to reproduce results
  - Growing maintenance burden
- Risk: CREDIT Defaulting.

- What to do?
  - Declare bankruptcy (aka switch to Anemoi)?
  - **No, restructure our investing strategy!**



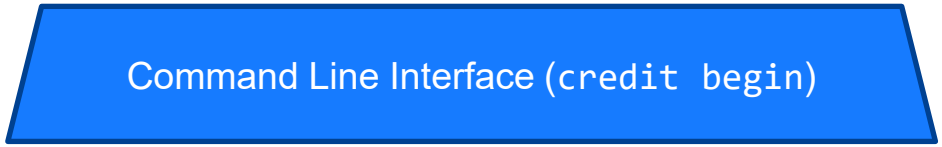
CREDIT Generation 2

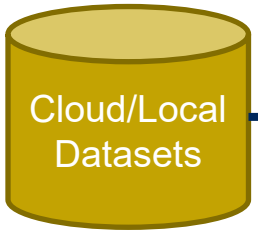
CREDIT: Community Research Earth Digital Intelligence Twin

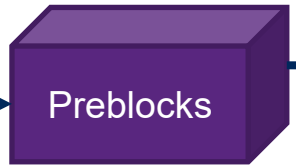


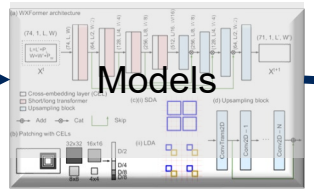
What is CREDIT?

An open, Python-based, foundational platform for developing **AI Earth System Emulators** [Schreck, et al. 2025].

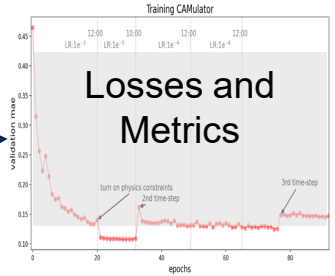






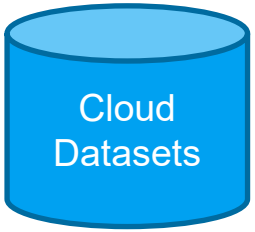








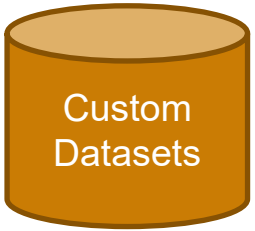
Datasets



Stream directly from ERA5, GOES, MRMS, GFS, GEFS, HRRR, or any other remote data archive.
Pros: Minimal local storage; **Cons**: latency, need reliable internet



Load from local collection of netCDF or Zarr files with consistent file paths
Pros: Low latency, flexible **Cons**: Need storage, data coverage



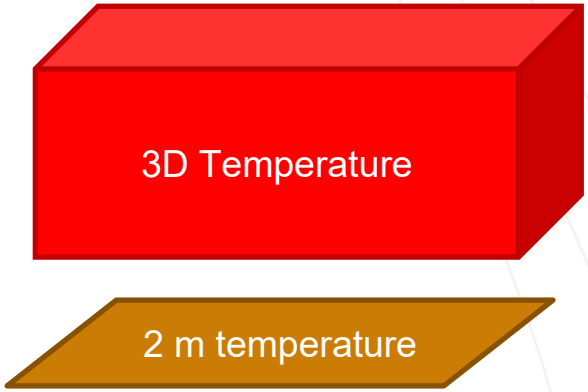
Can make custom PyTorch datasets that load from disk or generate synthetic or simulated data, like solar irradiance
Pros: Low latency, flexible **Cons**: may require more compute

State Dictionary

Sources

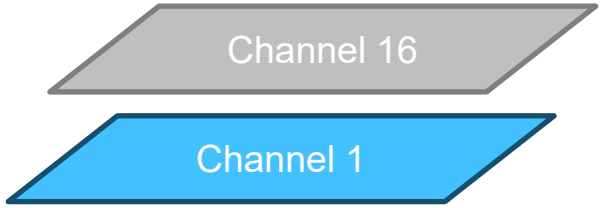
ERA5

Variable Types
Prognostic
Diagnostic
Dynamic Forcing
Static
Computed



GOES

Variable Types
Prognostic
Diagnostic
Dynamic Forcing
Static
Computed



Preblocks



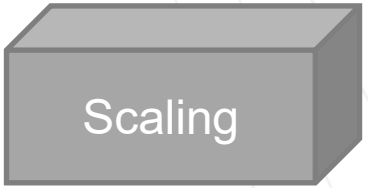
Making data variations more linear and Gaussian speeds up model convergence, especially data that varies exponentially across many orders of magnitude (e.g., microphysics).

CREDIT has log and sqrt transform blocks and can fill missing values.



Data from different sources should align on the same grid and vertical levels. Regrid applies transform weights from ESMF

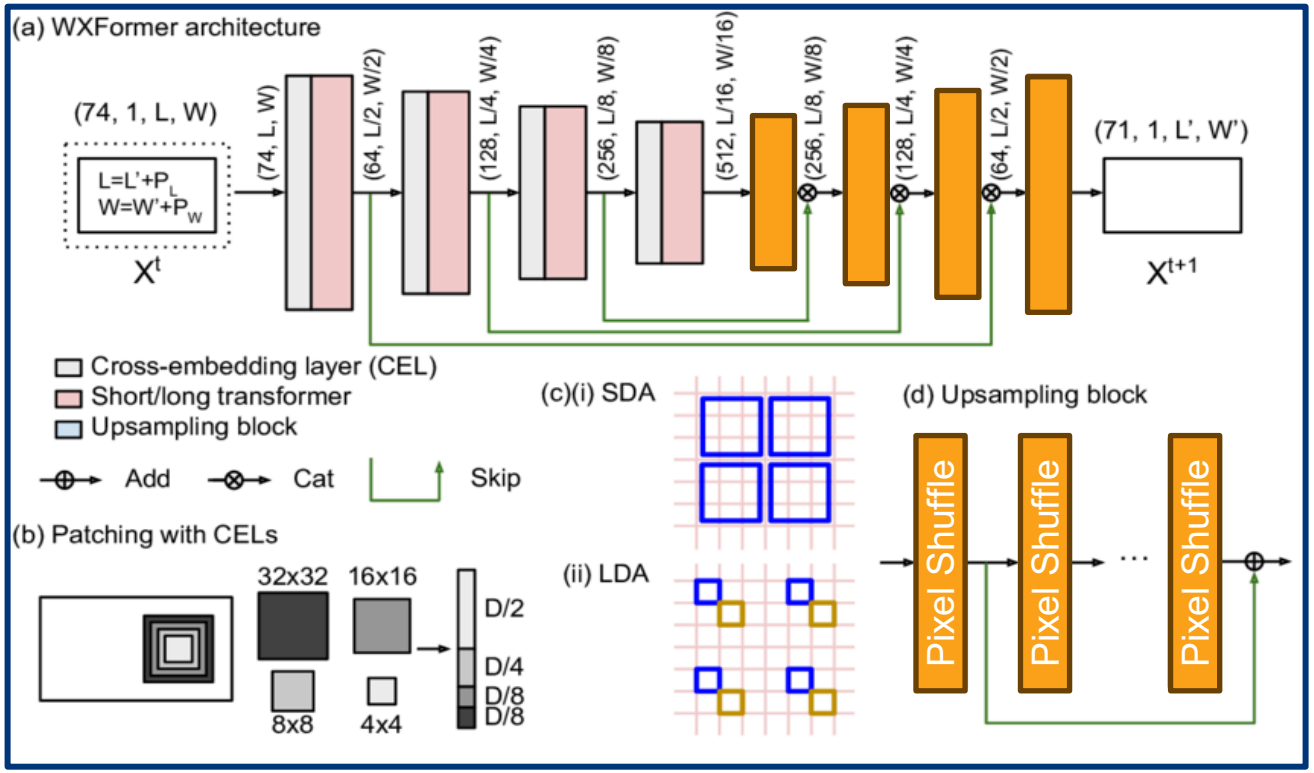
Vertical interpolate goes between different vertical coordinate systems.

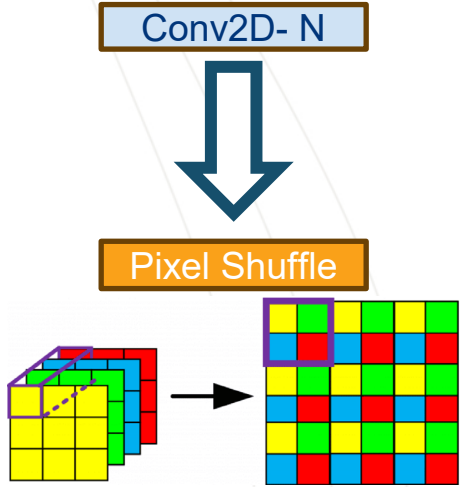


Data needs to be normalized after transforms to make optimization generally feasible.

CREDIT supports standard, minmax, and quantile scaling with Bridgescaler.

Models





Consequence:
Inherently drops decoder attention layers

Postblocks



Perform inverse transforms and scaling on predicted and target fields.

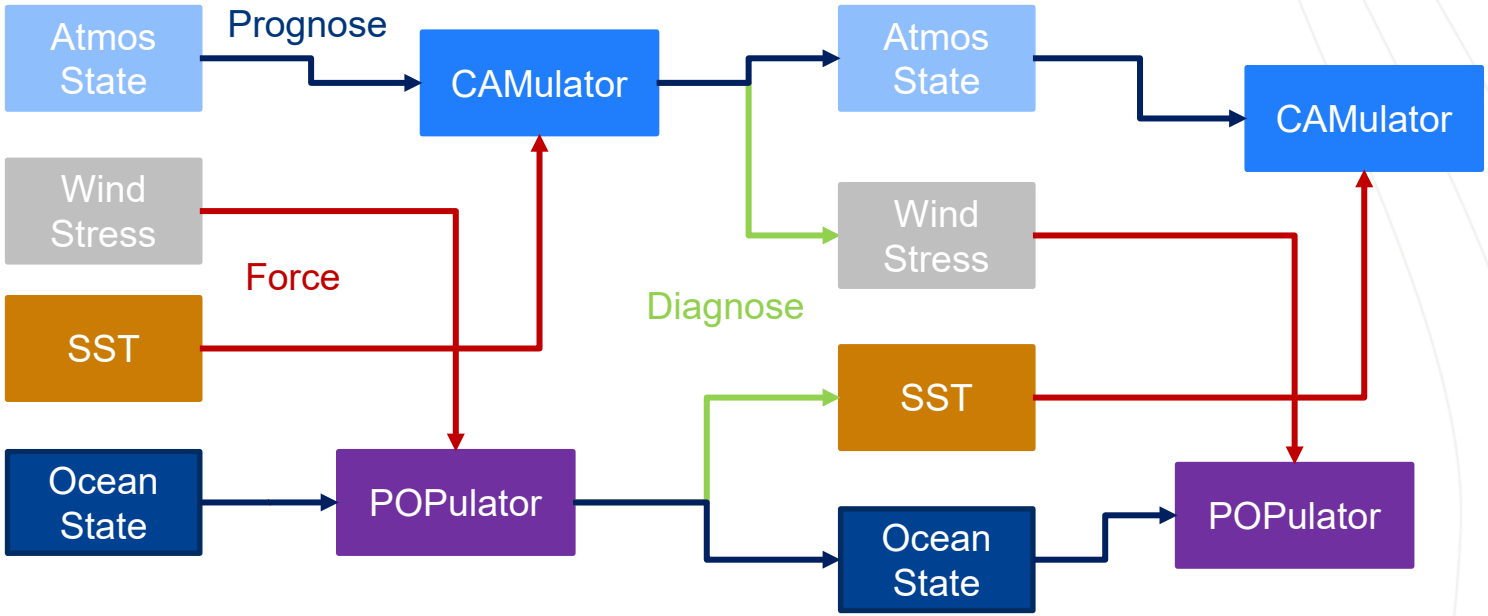


Calculate diagnostics like MSLP, geopotential, precipitable water, etc. from predicted state variables.
Guarantees exact relationships and fast and differentiable in PyTorch.



Enforce global mass, moisture budget, and energy conservation.

A Composable, Coupled Pipeline



Command Line Interface

```
$ credit -h
CREDIT — AI-NWP model training and inference platform

positional arguments:
  COMMAND
    begin           Interactively create a starter CREDIT config
    check           Validate a config: resolve every type, block, and path it names
    preprocess      Fit preprocessing scalers (BridgeScaler) over the training data
    train           Train a CREDIT v2 model
    rollout         Batch forecast rollout to NetCDF
    realtime        Operational realtime forecast (single init time)
    submit          Generate and submit a PBS or SLURM training, rollout, or realtime job
    plot            Quick global map: truth vs prediction from a saved checkpoint
    ask             Ask the CREDIT AI assistant a question

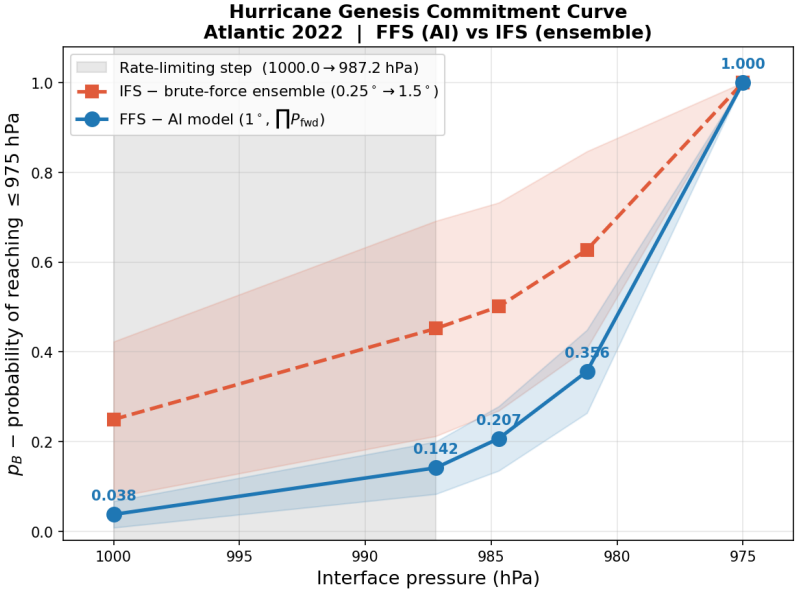
Examples:
  credit init       --grid 0.25deg -o my_config.yml
  credit preprocess -c config.yml
  credit train      -c config.yml
  credit realtime -c config.yml --init-time 2024-01-15T00 --steps 40
  credit rollout  -c config.yml
  credit submit   --cluster casper  -c config.yml --gpus 1
  credit submit   --cluster casper  -c config.yml --mode rollout --jobs 10
  credit submit   --cluster casper  -c config.yml --mode realtime --init-time 2024-01-15T00
  credit submit   --cluster derecho -c config.yml --gpus 4 --nodes 2
```

CREDIT Use Cases

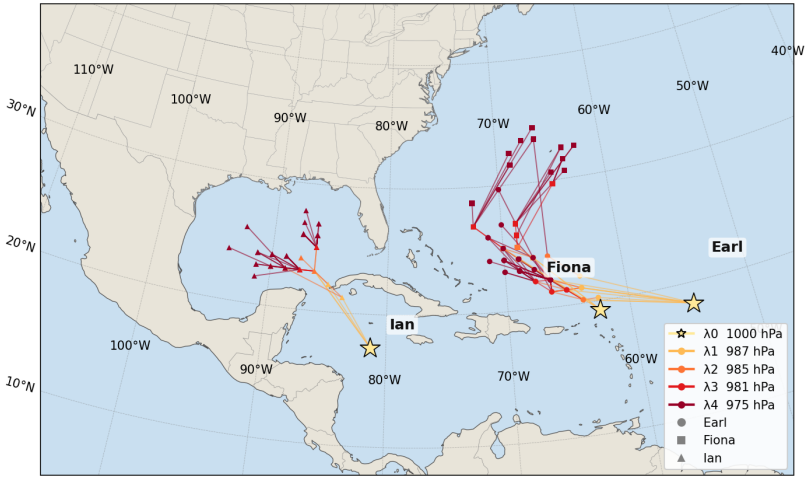
Rare Events Realized with Forward Flux Sampling

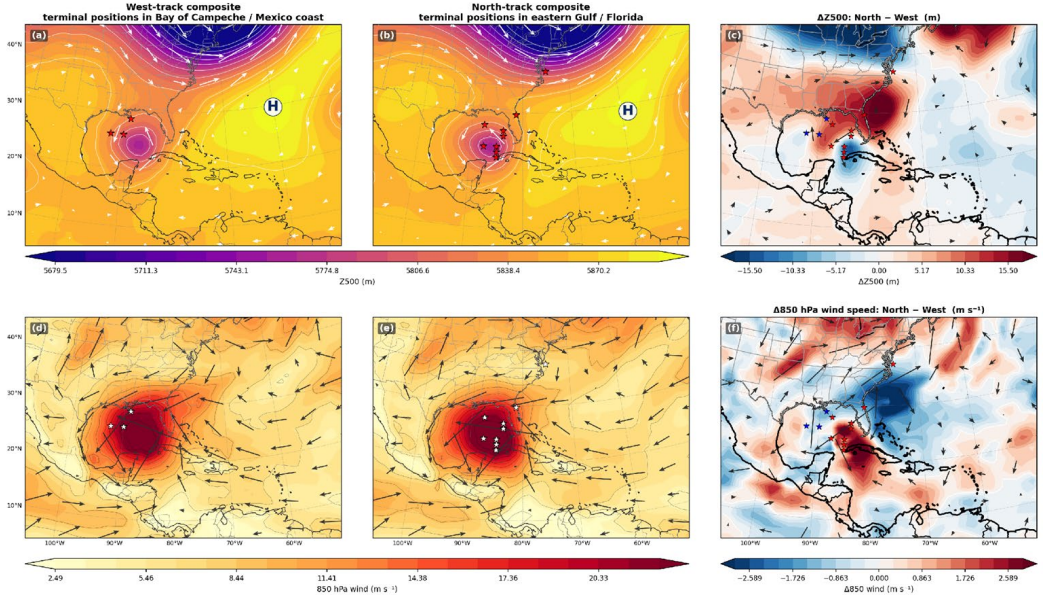
Calibrated stochastic perturbations to WXFormer enable re-sampling trajectories from states in individual runs. Preferential sampling toward a target give us a plausible extreme trajectory and calibrated probabilities of tail events.

No fine-tuning or guided latent perturbation needed!

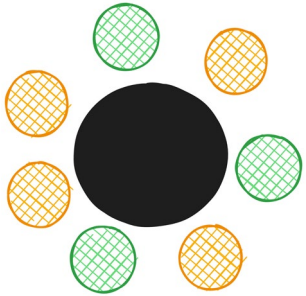


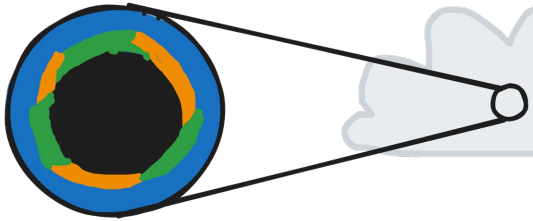
Forward Flux Sampling of Hurricane Intensity

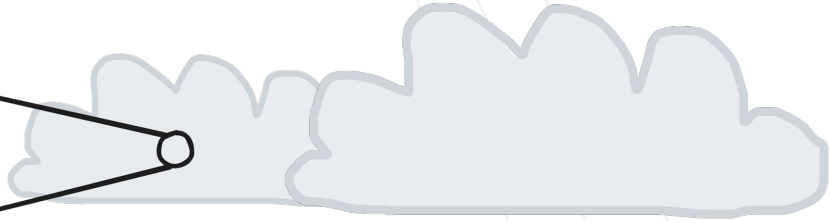


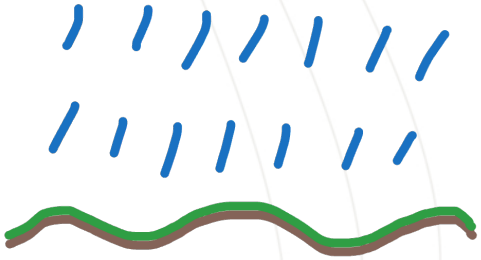


CREDIT- chem for Black Carbon











Black Carbon (BC) comes from incomplete combution of carbon-based fuels, like wood or diesel

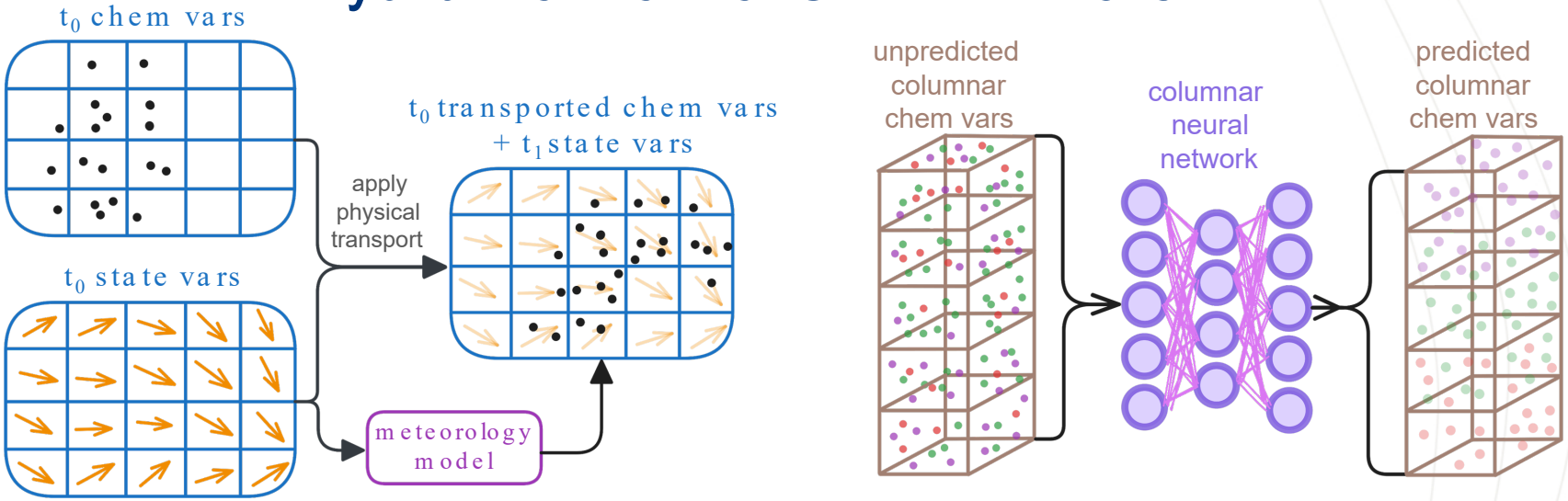
Black Carbon (BC)

Simulating primary organic matter (POM), secondary organic aerosols (SOA), and sulfates as well

CREDIT Chem Team
LJ Dunphy, Alma Roux, Ben Gaubert, DJ Gagne, John Schreck, Charlie Becker

CREDIT- chem Model Setup

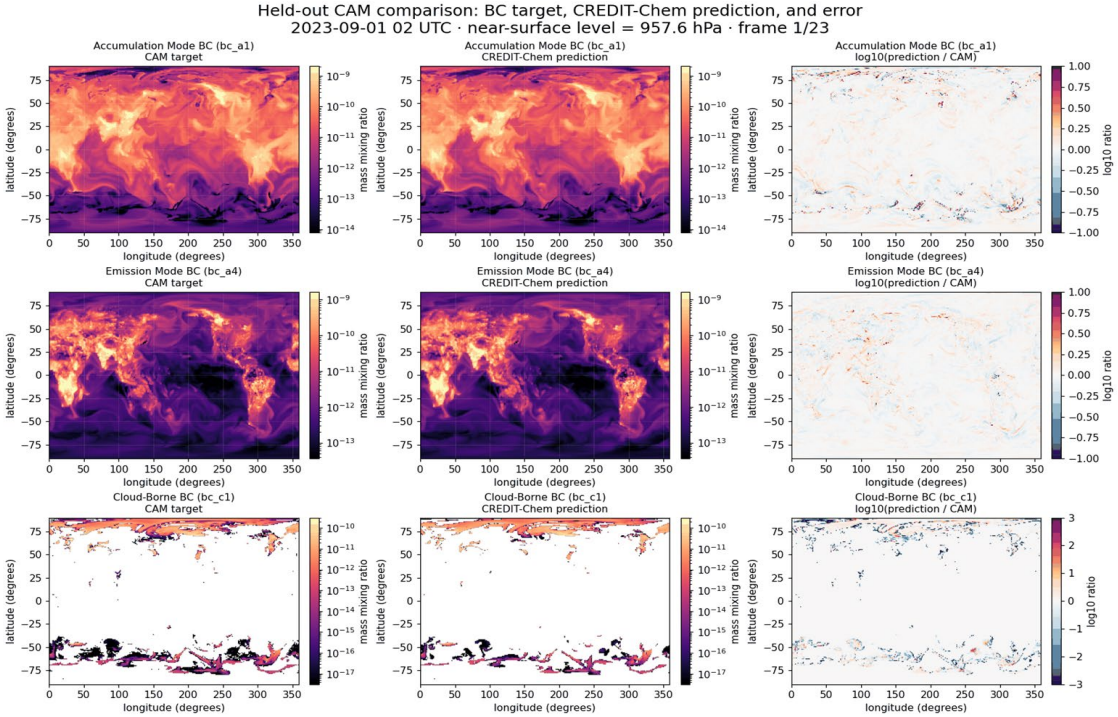
Hybrid Workflow for CREDIT - chem

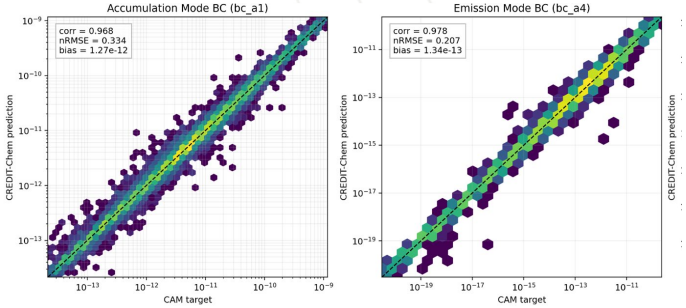


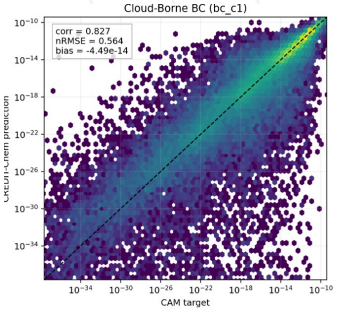
NOTE: Physical transport is NOT ML based

Trained on CAM6 - chem, 1 hour time step,
~1 deg. grid spacing, 32 vertical levels.
Predicts: Black Carbon (BC), Primary Organic Matter (POM), Secondary Organic Aerosols (SOAs), and Sulfates

CREDIT- chem Results





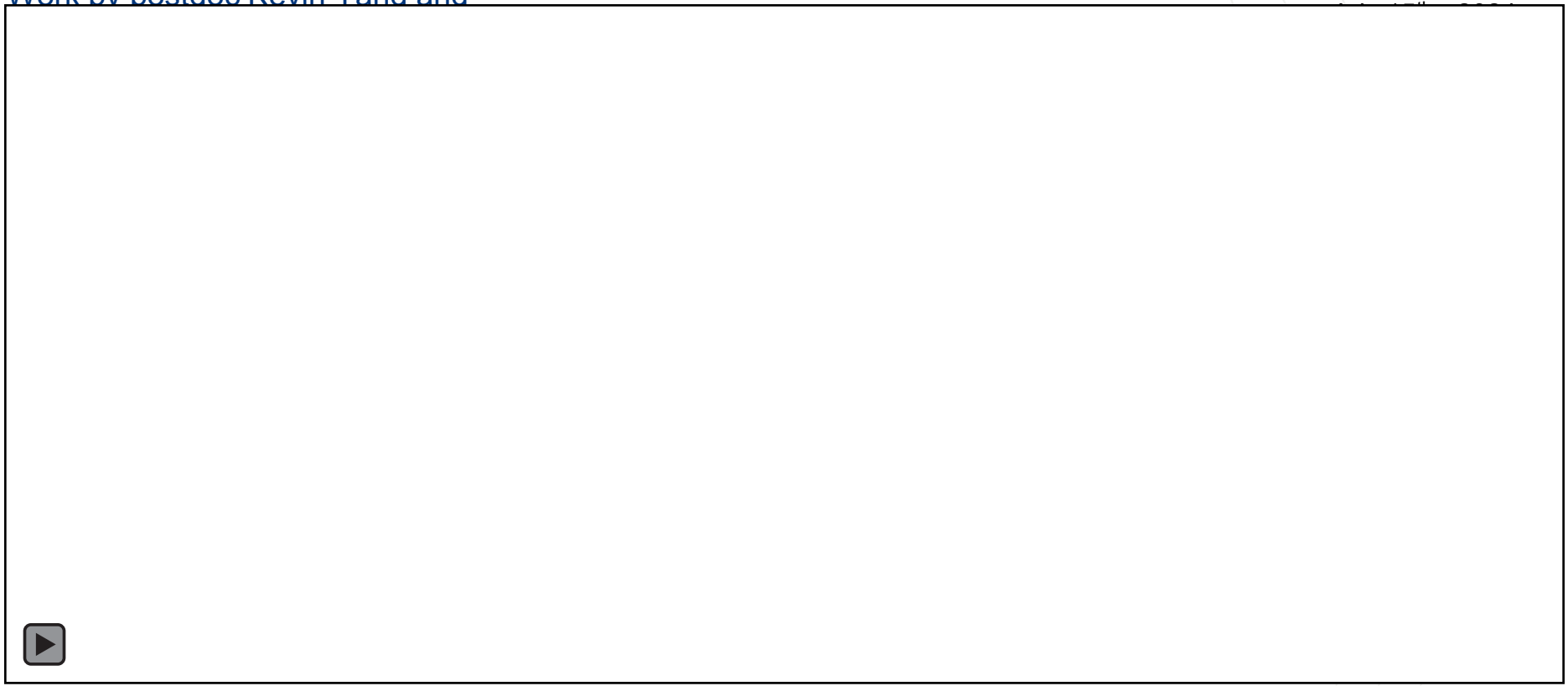


CREDIT- Chem captures evolution of each stage of black carbon.

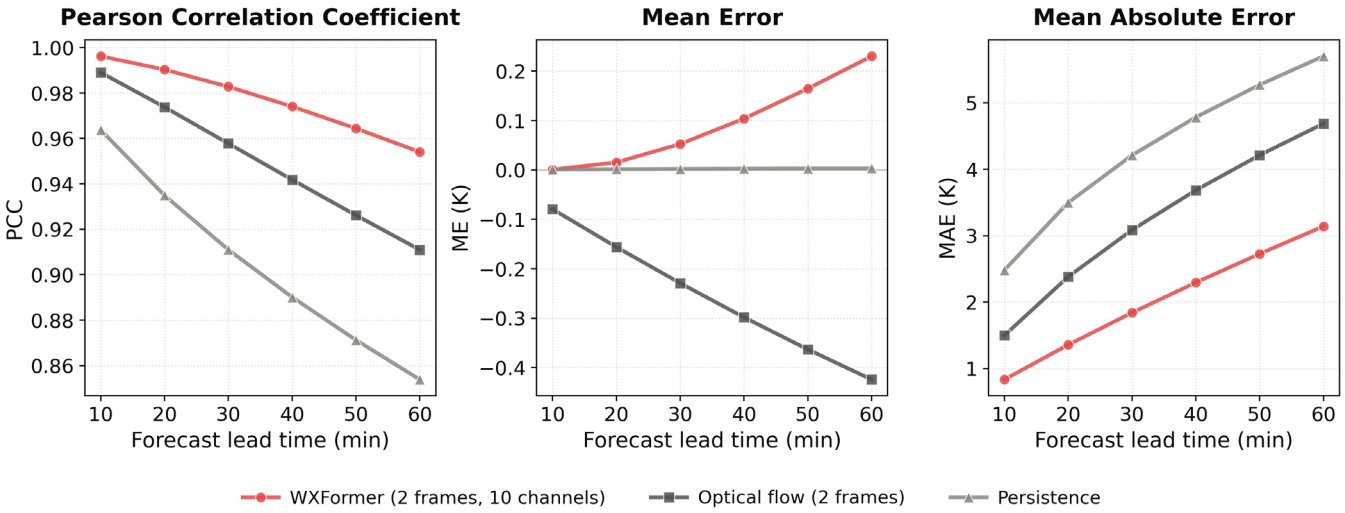
Can produce stable rollouts out a month at hourly time steps.

GOES 10.3 μm Nowcasting over CONUS

Work by postdoc Kevin Yang and



Enhancing Storm Growth Prediction with Temporal Context



CAMulator : Now Open Source and Weights

willychap / **camulator** Like 0

Other | PyTorch | miles-credit | climate | weather | atmosphere | emulator | eart

Model card | Files and versions xet | Community

CAMulator

CAMulator is an auto-regressive machine-learned emulator of NSF NCAR's CAM6 atmosphere, trained and run within the CREDIT framework. Given prescribed sea-surface temperature, sea-ice, incoming solar radiation, and CO2, it rolls a 1 degree (192x288), 32-level, 6-hourly atmospheric state forward for climate-length simulations (years to decades). It conserves global dry-air mass, moisture, and total atmospheric energy, remains numerically stable over decadal rollouts, and reproduces the annual climatology together with major modes of variability such as ENSO and the NAO -- at roughly a 350x speedup over CAM6, making it an efficient way to generate large climate ensembles.

The model and method are described in Chapman et al. (2025), *CAMulator: Fast Emulation of the Community Atmosphere Model* (arXiv:2504.06007).

CAMulator is a research tool. It emulates a specific CAM6 configuration and is not a substitute for an operational forecast or a full Earth-system model.

NVIDIA | Earth2Studio | main

CAMulator

Global | CM | 2025 | 24 GB | NCAR | PyTorch

Import path: `earth2studio.models.px.CAMulator`

View source on GitHub → | View install commands →

Documentation

Bases: `Module`, `AutoModelMixin`, `PrognosticMixin`

CAMulator: NSF NCAR's autoregressive emulator of the CAM6 atmosphere, trained in the CREDIT framework. A CrossFormer encoder-decoder advances a 1 degree (192x288), 32 hybrid-sigma-level atmospheric state by 6 hours given prescribed sea-surface temperature, sea-ice fraction, TOA insolation and CO2 forcing, for climate-length rollouts. The inference-time post-processing of the CREDIT toolbox is reproduced: tracer clipping, the jet wind-artifact filter and the global dry-air mass, water and total-energy conservation fixers.

Announcements

NSF NCAR is and will be hiring!

- ASP Postdoc (October 31)
- SIParCS interns (undergrad and grad, due late Dec)
- Postdoc with Chris Snyder
- ML Engineers
- Science positions

Save the Date!
Emulated and Hybrid ML Physics Earth System Models Workshop
May 12- 14, 2027
Mesa Lab
Boulder, Colorado

Takeaways

- CREDIT Gen 2 provides the scaffolding to build coupled AI Earth System models.
- The speed of AI Earth system emulators enables new interactive applications previously impossible with traditional models.
- Modular components for the whole pipeline enable mechanistic interpretability

CREDIT Source: `github.com/NCAR/miles-credit`



Contact Me
dgagne@ucar.edu

Thank you to the CREDIT contributors, researchers, and supporters!

MILES: John Schreck, David John Gagne, Charlie Becker, Gabrielle Gantos, Kevin Yang

IMPACT: Katelyn FitzGerald, Scott Pearse, Nihanth Cherukuru, Matt Rehme, Stas Jaroszynski, Kate Kuzemka, John Clyne

CISL: Negin Sobhani, Ben Kirk, Daniel Howard, Doug Schuster, Jeff Anderson

Rest of NCAR: Yingkai "Kyle" Sha, Dhamma Kimpara, Kirsten Mayer, Judith Berner, Dan Amhrein, Matt Haymann, Ethan Gutmann, Pedro Jimenez, Andy Newman, Sue Haupt, Chris Snyder, Alma Roux, Scott Swerdlin, Chris Castro

University Community: Will Chapman, Arnold Kazadi, Stanley Akor, Aidan Janney, Laure Zanna, Amy McGovern, Stuart Edris, Reyna Wu, Aneesh Subramanian, Gio Ellis, Steve Greybush, Christian Spallone, Daniel Neamati, LJ Dunphy, Ethan Campbell