Gaia: Global AI Accelerator:

Modeling MJO structures and tipping point analysis

Contractor's Name and Address:
Systems & Technology Research
600 West Cummings Park
Woburn MA 01801

UNSW
Climate Change
Research Centre

Milestone 5:
Potential datasets (and providers) for Phase 2 to address predictability of climate effects at 1-to-3 decade time scales and regional or global spatial extents.

Date of Report:

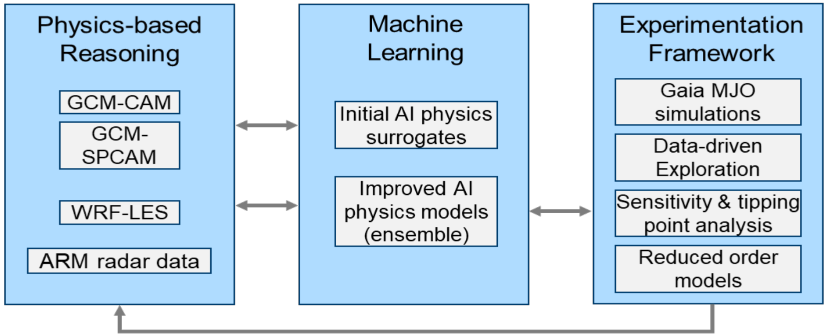
June 13, 2022

Gaia: Global AI Accelerator

UNSW
Climate Change
Research Centre

What's hard?

- GCMs are computationally expensive and lack the resolution needed to adequately model local convection and thus clouds
- This greatly increases GCM forecast errors and impedes propagation of large-scale wave phenomena such as the MJO



What will GAIA accomplish?

- Enable a GCM to accurately model local convection and predict self-organizing atmospheric wave phenomena (i.e. improved fidelity with 5X – 10X speedup)
- Exploit this GCM to explore possible future regimes and identify early warning signatures for MJO-related tipping points.

Progress

- Optimized and validated high-skill AI surrogates for multiple GCM local models
- Investigated bottlenecked auto-encoder architectures to quantify intrinsic dimensionality of the cloud parameterization model
- ***Next Steps***: Increase dataset from 4 to 10 years simulation time; Integrate surrogates into MJO; Improve surrogates using WRF-LES and observational data; study onset of tipping points using reduced order models

M5 Update

Description of datasets, data preprocessing, python modeling & analytics, and data products can be found at:

https://github.com/stresearch/gaia

Training of initial AI cloud physics model surrogates is currently based on 3 and 4-year runs from two NCAR community atmospheric models (CAM4 and SPCAM), now being extended out to 10 years of simulation time:

GCM

- Community Atmospheric Model (CAM4)
- 30 minute time-step
- 2.5-degree grid (144x96)
- 30 altitude levels
- Four year run (1979 SST; Time Varying) which will be extended to ten years.
- Outputs every 3 hours + additional model time-step (memory)

CRM

- SPCAM (super parameterized CAM)
- 20 minute time-step
- 16 SAM (The System for Atmospheric Modeling) Columns
- 26 levels
- Year 2000 SST (Climatology)
- Three year simulations:
  - Morrison Microphysics + Conventional parameterization for moist convection and large-scale condensation.
  - Morrison Microphysics + Higher-order turbulence closure scheme, Cloud Layers Unified By Binormals (CLUBB)
- Outputs every 3 hours + additional model time-step (memory)

Gaia datasets

UNSW
Climate Change
Research Centre

Surrogate Inputs

Name	Long Name	shape	unit
Q	Specific humidity	(T, L, 96, 144)	kg/kg
T	Temperature	(T, L, 96, 144)	K
U	Zonal wind	(T, L, 96, 144)	m/s
V	Meridional wind	(T, L, 96, 144)	m/s
OMEGA	Vertical velocity (pressure)	(T, L, 96, 144)	Pa/s
PSL	Sea level pressure	(T, 96, 144)	Pa
SOLIN	Solar insolation	(T, 96, 144)	W/m2
SHFLX	Surface sensible heat flux	(T, 96, 144)	W/m2
LHFLX	Surface latent heat flux	(T, 96, 144)	W/m2
FSNS	Net solar flux at surface	(T, 96, 144)	W/m2
FLNS	Net longwave flux at surface	(T, 96, 144)	W/m2
FSNT	Net solar flux at top of model	(T, 96, 144)	W/m2
FLNT	Net longwave flux at top of model	(T, 96, 144)	W/m2
Z3	Geopotential Height (above sea level)	(T, L, 96, 144)	m

Gaia datasets

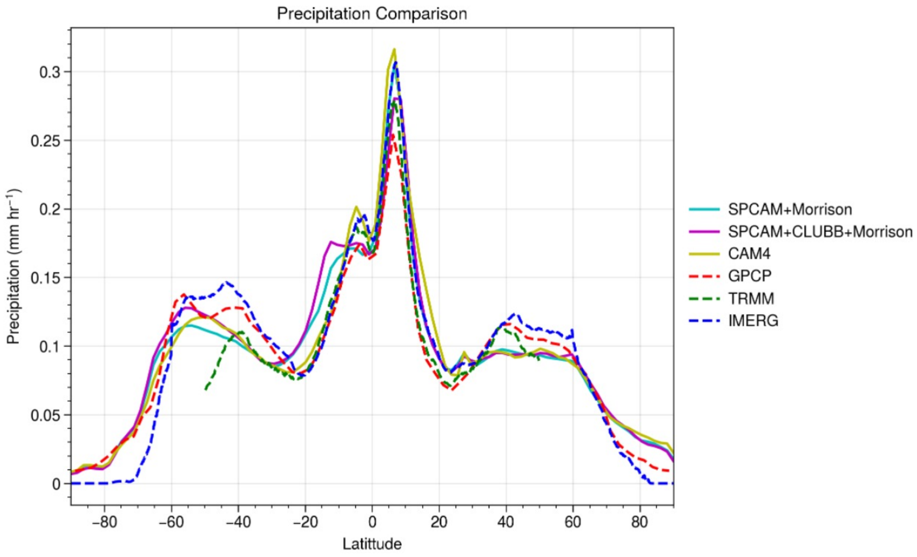
Surrogate Outputs

Name	Long Name	shape	unit
PRECT	Total (convective and large-scale) precipitation rate (liq + ice)	(T, 96, 144)	m/s
PRECC	Convective precipitation rate (liq + ice)	(T, 96, 144)	m/s
PTEQ	Q total physics tendency	(T, L, 96, 144)	kg/kg/s
PTTEND	T total physics tendency	(T, L, 96, 144)	K/s

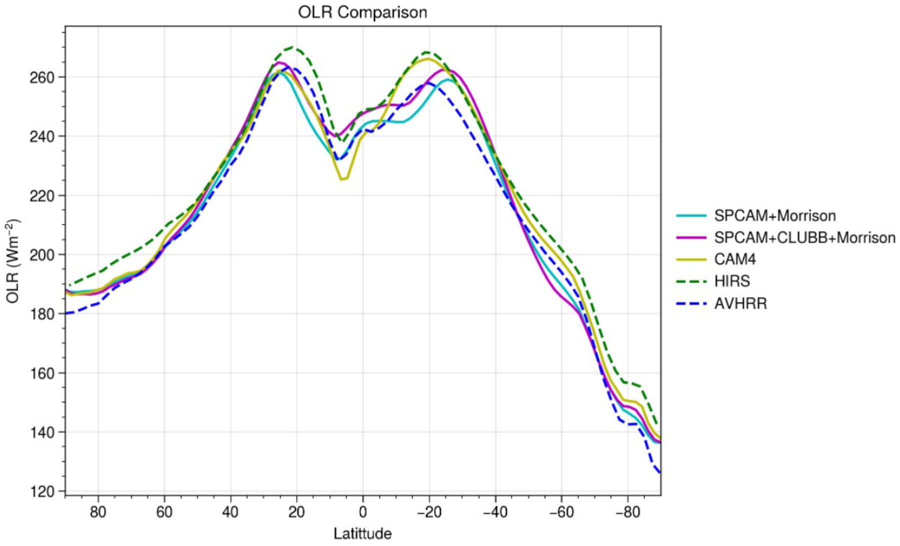
Gaia datasets

All the simulation datasets have been evaluated zonally for Precipitation and Outgoing Longwave Radiation (OLR), comparing to satellite observations

UNSW
Climate Change
Research Centre



Zonal distribution of precipitation using model runs (solid lines) and satellite observations (dashed lines)



Zonal distribution of Outgoing Longwave Radiation (OLR) using model runs (solid lines) and satellite observations (dashed lines)

Gaia datasets

UNSW
Climate Change
Research Centre

Having trained different AI surrogates in CAM4 or SPCAM datasets, we evaluate how well CAM4-trained are predictive of SPCAM simulations and vice versa

Areas of high discrepancy will help identify areas for retraining based on the NCAR WRF model. These new training data will be used to learn corrections to the CAM4 and SPCAM-trained surrogates.

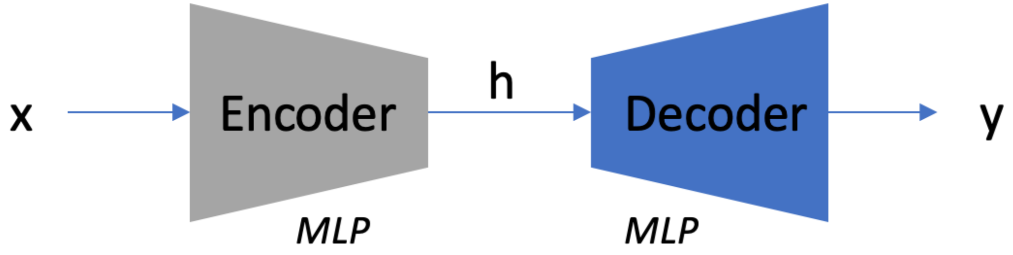
- WRF (Weather Research and Forecasting Model)
- 50 km x 50 km domains; periodic boundary conditions
- 100 levels
- 100 weather cases
- 1 week spin up at 2 km + 5 day simulation at 200m resolution (LES)
- 3 hourly Boundary Conditions by SPCAM runs + nudging of state variables
- History outputs at 10 minutes (horizontally averaged and mapped to same vertical grid as CAM4)

New (for M5) data analyses

Our upcoming tipping point analyses will make heavy use of reduced-order models. In order to gauge the intrinsic dimensionality of the input data to the AI surrogate, we constructed another AI network model based off of autoencoder architectures

The encoder block maps the input vectors **x** to a bottlenecked latent representation **h**

The decoder block maps the latent representation **h** to the output vector **y**



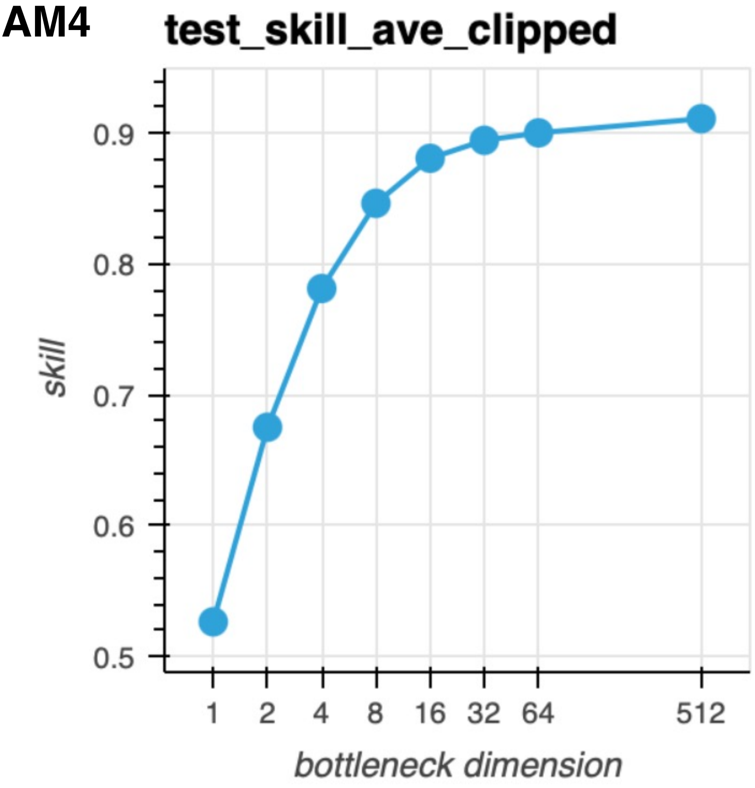
New (for M5) data analyses

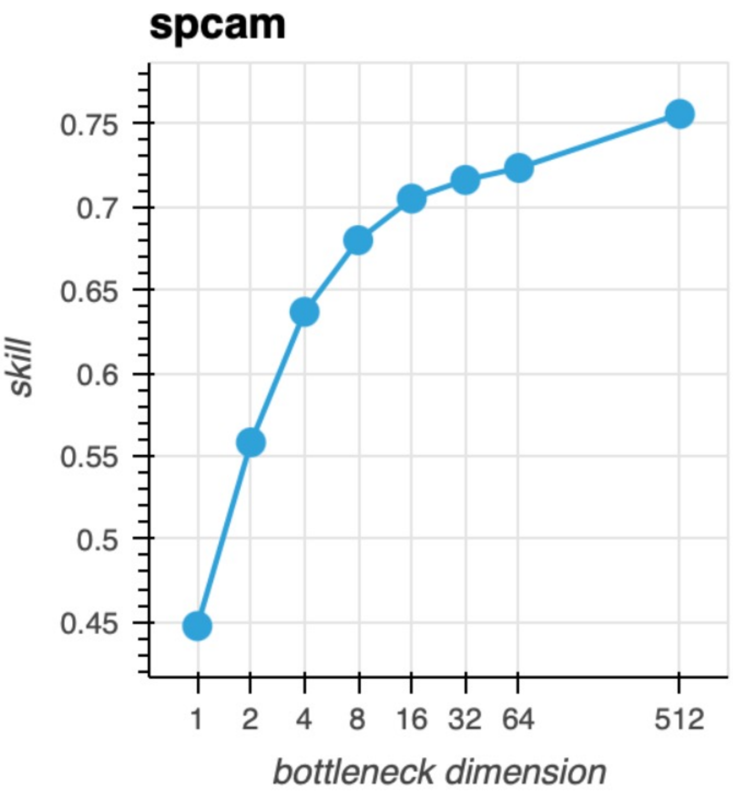
UNSW
Climate Change
Research Centre

Skill for CAM4 and SPCAM using bottlenecked encoder-decoder AI models

Skill begins to saturate at a dimension of ~ 32

CAM4





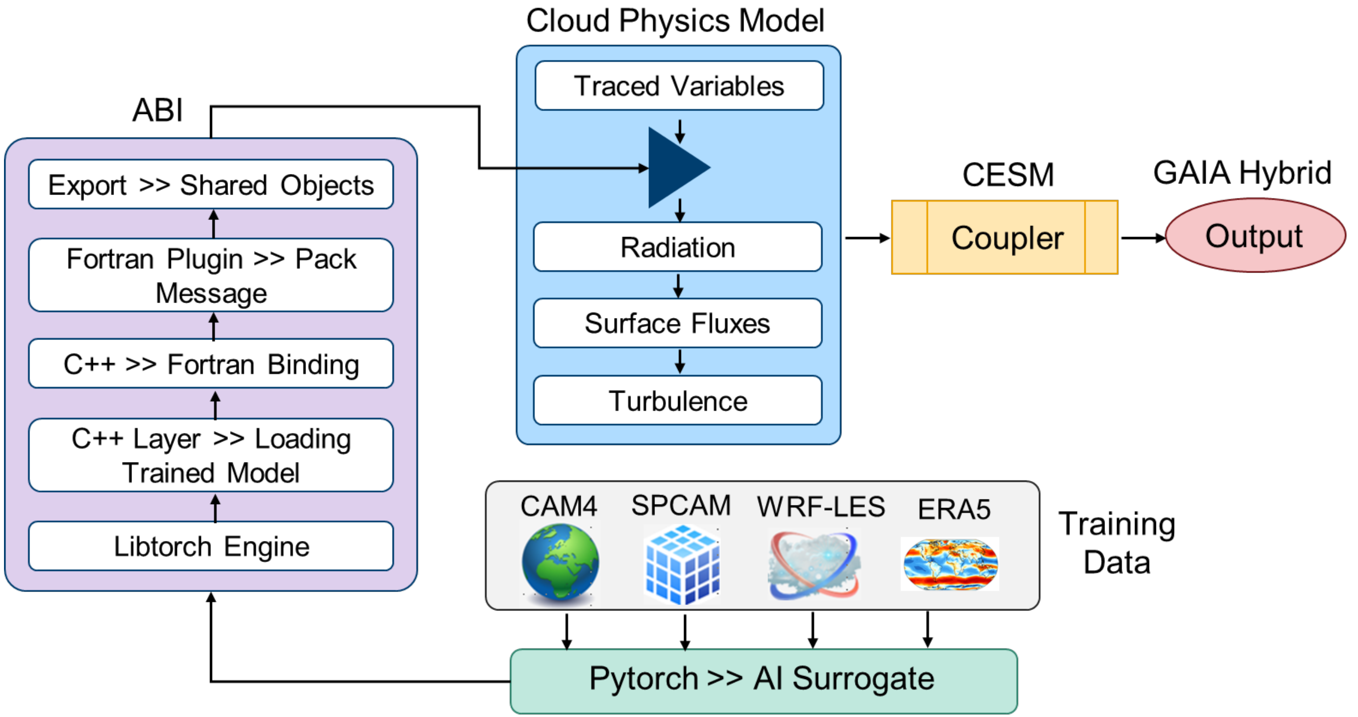
Key next step: hybrid model integration

We are still working on integrating the pytorch-based AI surrogate back into the Fortran-based GCM models

Once integrated, hybrid model stability will be tested for a diverse set of AI surrogates, with and without WRF- and ERA5-retraining

Application Binary Interface (ABI)

- Start with total T,Q physics tendencies and track them back
- Export Pytorch Model with Torchscript
- Bypass Python by using C++
- Call C++ within Fortran


Cloud Physics Model
Traced Variables
Radiation
Surface Fluxes
Turbulence
ABI
Export >> Shared Objects
Fortran Plugin >> Pack Message
C++ >> Fortran Binding
C++ Layer >> Loading Trained Model
Libtorch Engine
CESM
Coupler
GAIA Hybrid
Output
CAM4
SPCAM
WRF-LES
ERA5
Training Data
Pytorch >> AI Surrogate