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 4:

Beta version of surrogate models, including performance analyses and identified areas for model improvement

Date of Report:

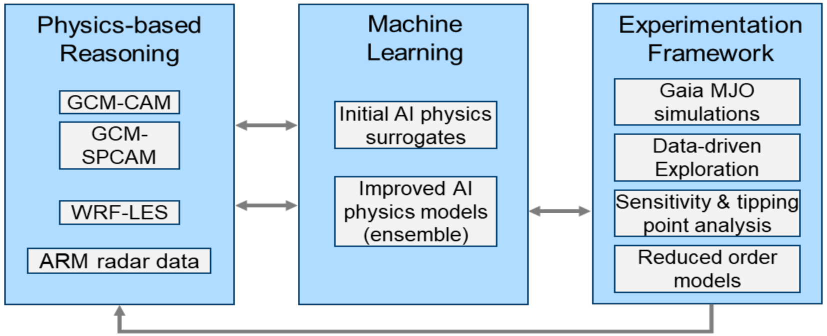
May 19, 2022

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



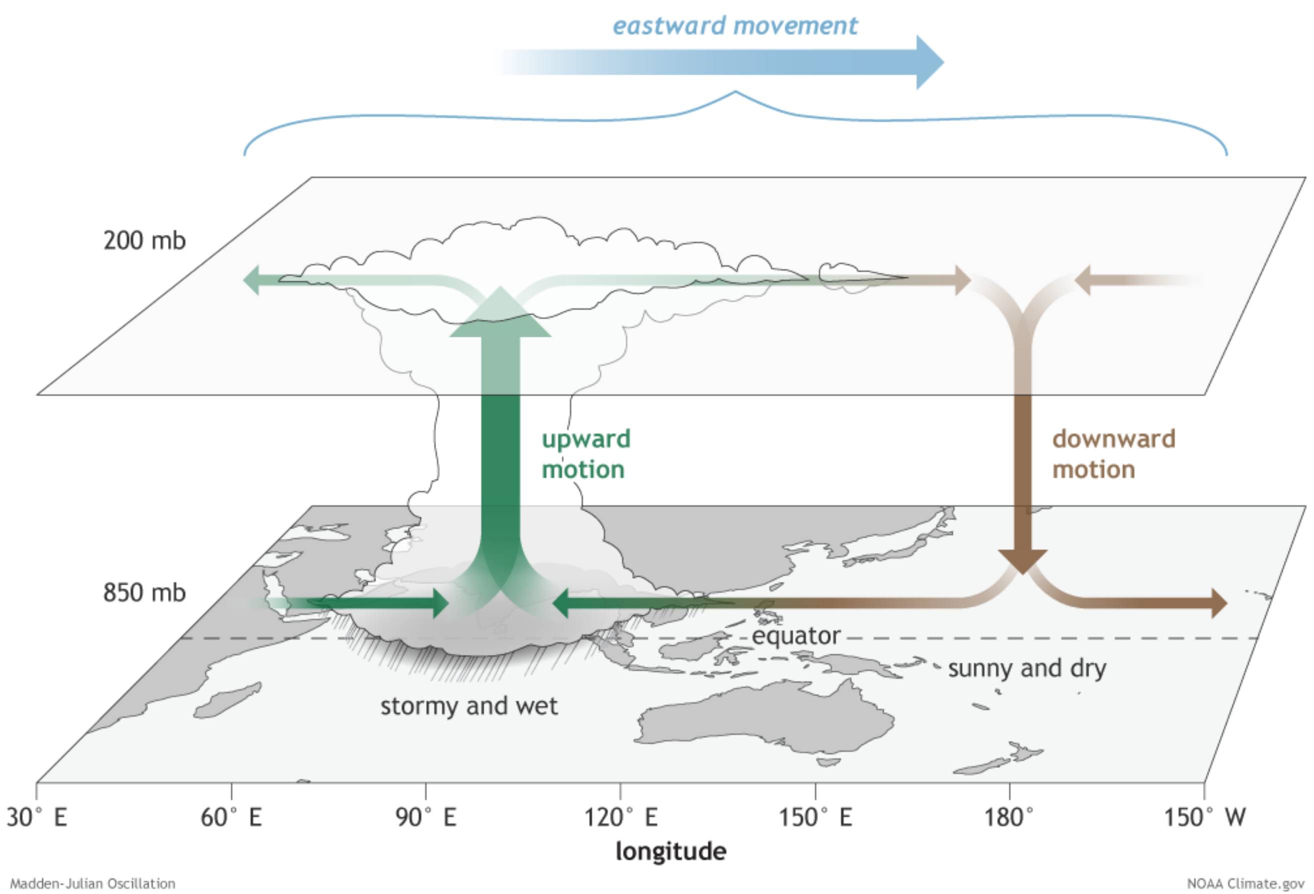
What will GAIA accomplish?

- Enable a GCM to accurately model local convection and predict self-organizing atmospheric wave phenomena
- 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
- Studied parametric dependencies and benefits of additional memory terms
- Next Steps: Integrate surrogates into MJO; Improve surrogates using WRF-LES and observational data; study onset of tipping points using reduced order models

Madden Julian Oscillation (MJO)



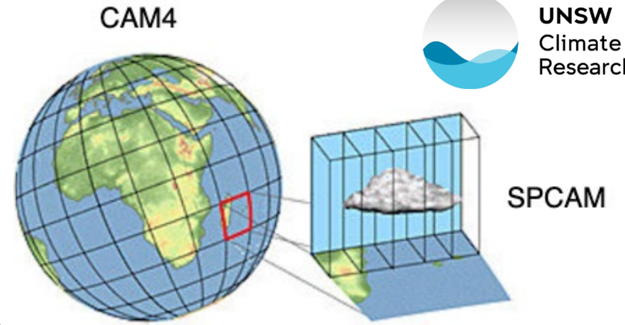
The MJO is an eastward moving organization of clouds, rainfall, winds, and pressure that traverses the planet in the tropics and returns to its initial starting point in 30 to 60 days, on average.

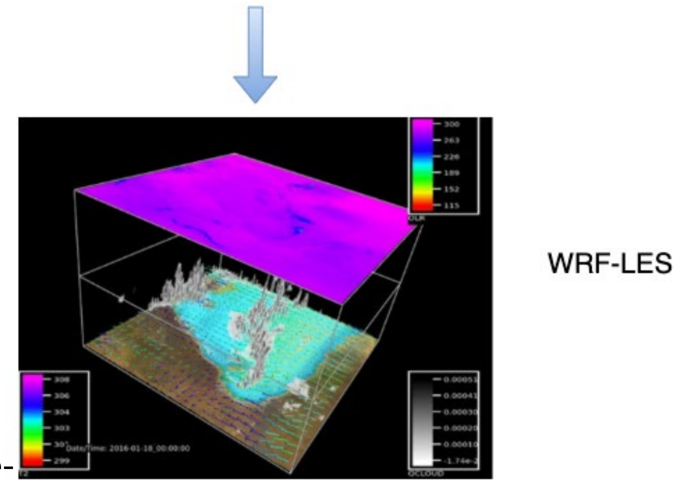
Could it intensify rapidly in a warmer climate?

Gaia Datasets

str

UNSW
Climate Change
Research Centre





GCM

- Community Atmospheric Model (CAM4)
  - 30 minute time-step – each time step output
  - 2.5-degree grid (144x96), 30 levels
  - Four-year run (time varying sea surface T (SST)), will be extended to ten years
  - Ten years at +4K SST (May/June)

CRM-GCM

- SPCAM (super parameterized CAM)
  - 30 minute time-step – each time step output
  - 16 SAM (The System for Atmospheric Modeling) Columns, 26 levels
  - Four years (fixed SST) - will do 10 years varying SST, +4K
  - Two versions:
    - Morrison Microphysics + Conventional parameterization for moist convection and large-scale condensation.
    - Morrison Microphysics + Higher-order turbulence closure scheme, Cloud Layers Unified By Binormals (CLUBB) – <u>Version we will go forward with</u>

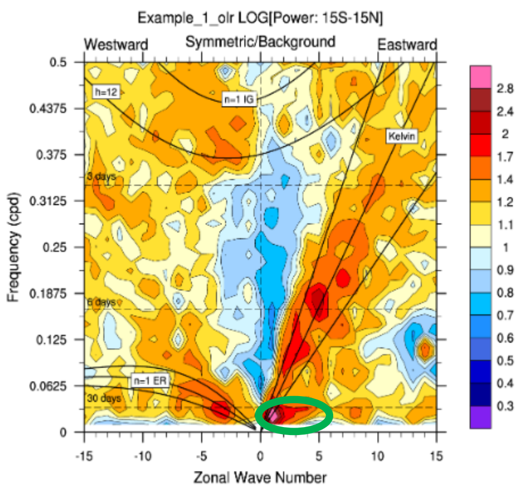
GAIA initial objective: build superior ML surrogate to replace CAM physics

Madden Julian Oscillation (MJO)

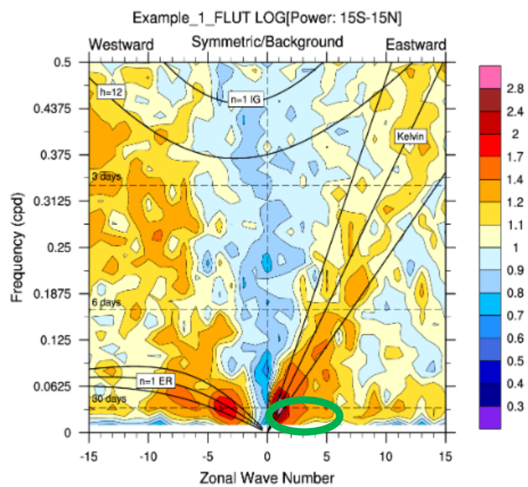
str

UNSW Climate Change Research Centre

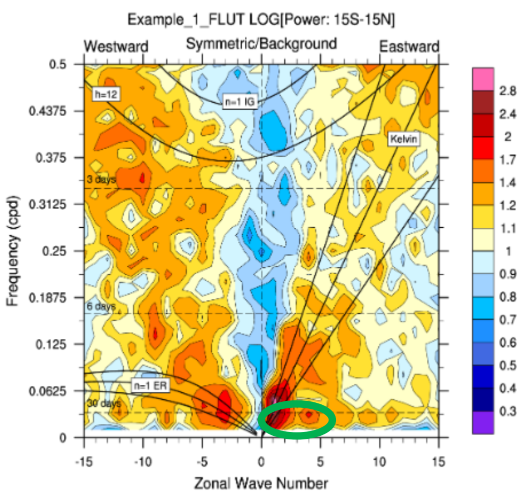
Observations



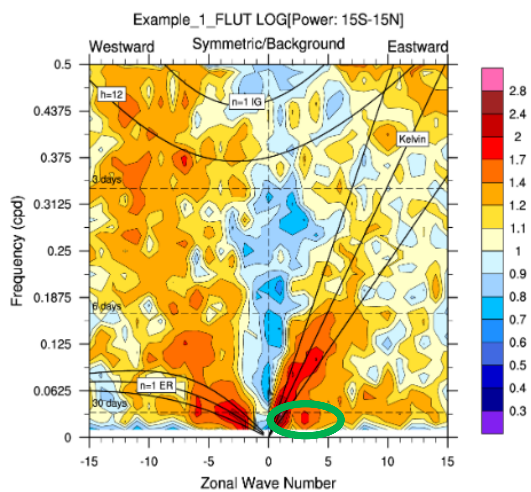
CAM4



SPCAM



SPCAM + CLUBB



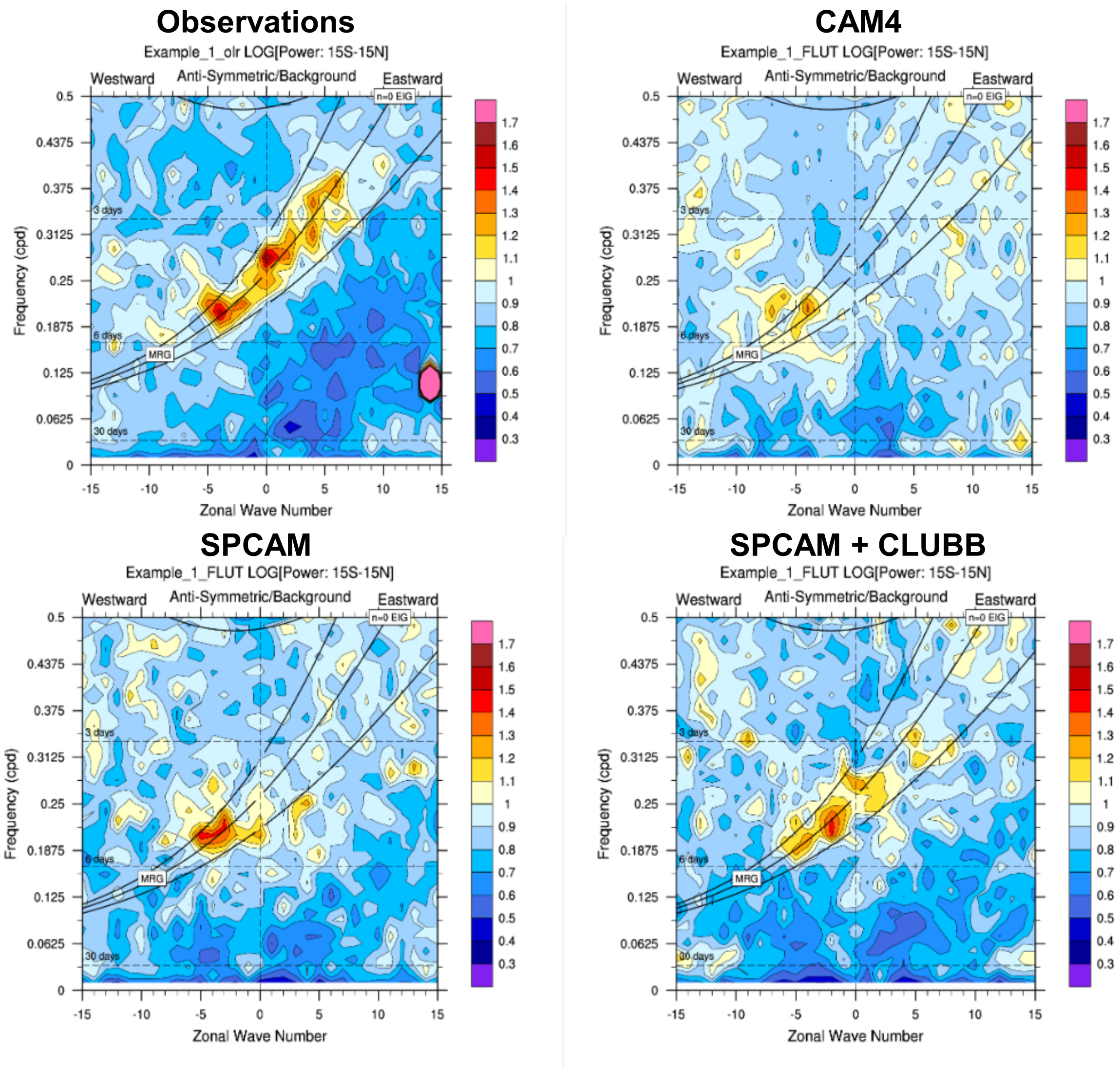
Symmetric wave structures (Northern/Southern hemisphere-symmetric)

Freq vs. Zone Wave Number

MJO is weak in all three GCMs

Kelvin wave activity is also very weak (better in SPCAM+CLUBB)

Madden Julian Oscillation (MJO)



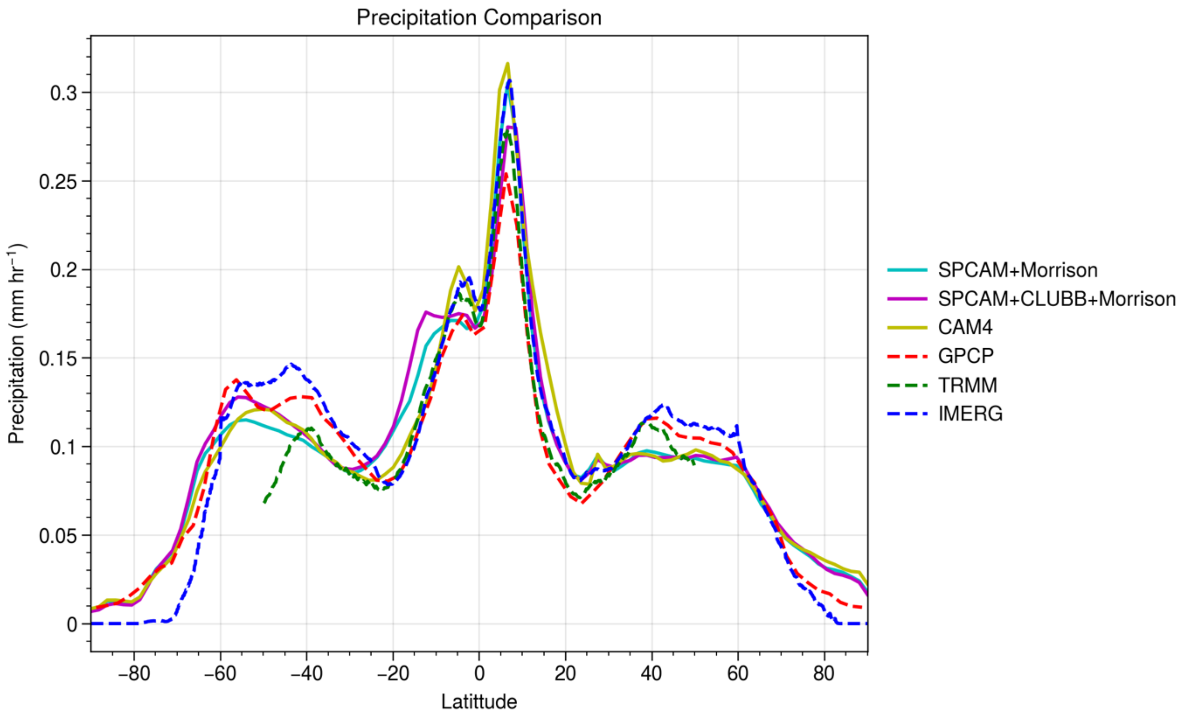
Anti-symmetric wave structures (Northern/Southern hemisphere-antisymmetric)

Freq vs. Zone Wave Number

Asymmetric Rossby wave activity is also too weak

GAIA GCM Evaluation, (1)

All simulations evaluated zonally for Precipitation and Outgoing Longwave Radiation (OLR) from satellite observations

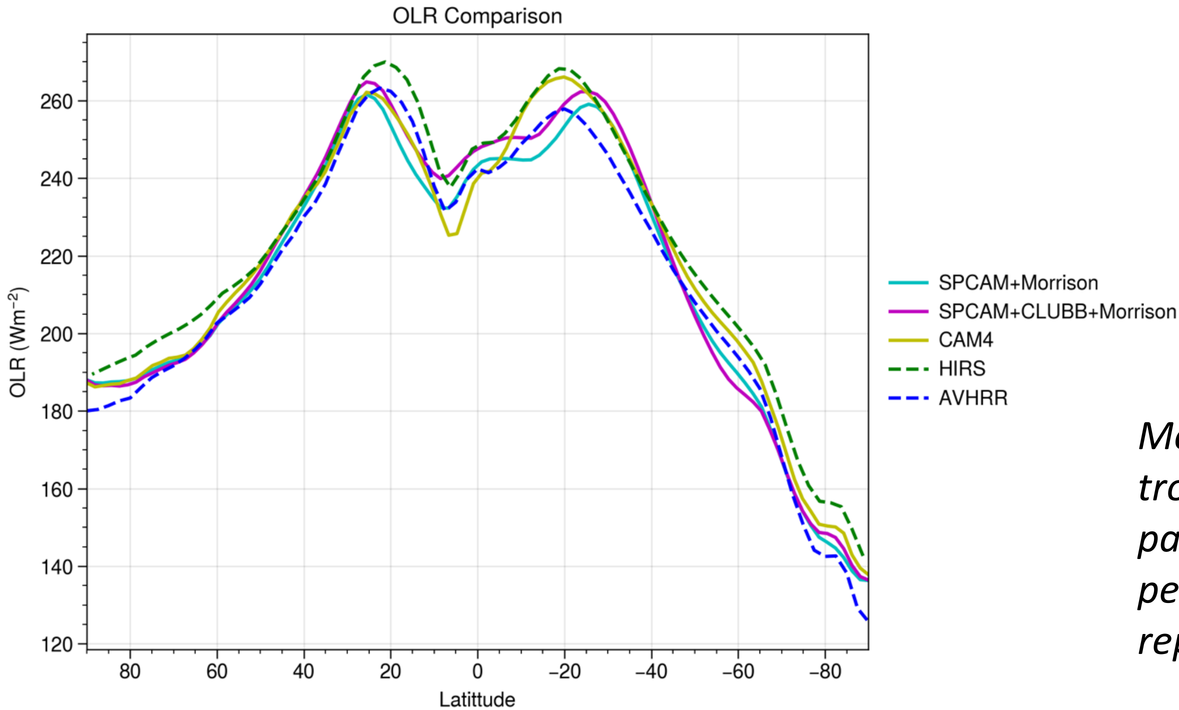


Zonal distribution of precipitation using model runs (solid lines) and satellite observations (dashed lines)

GAIA GCM Evaluation, (2)

str

UNSW Climate Change Research Centre



Modelled OLR is mostly underestimated in the tropics, but SPCAM with CLUBB (unified cloud parameterisation) and Morrison microphysics perform better, likely due to the representation of deeper clouds

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

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

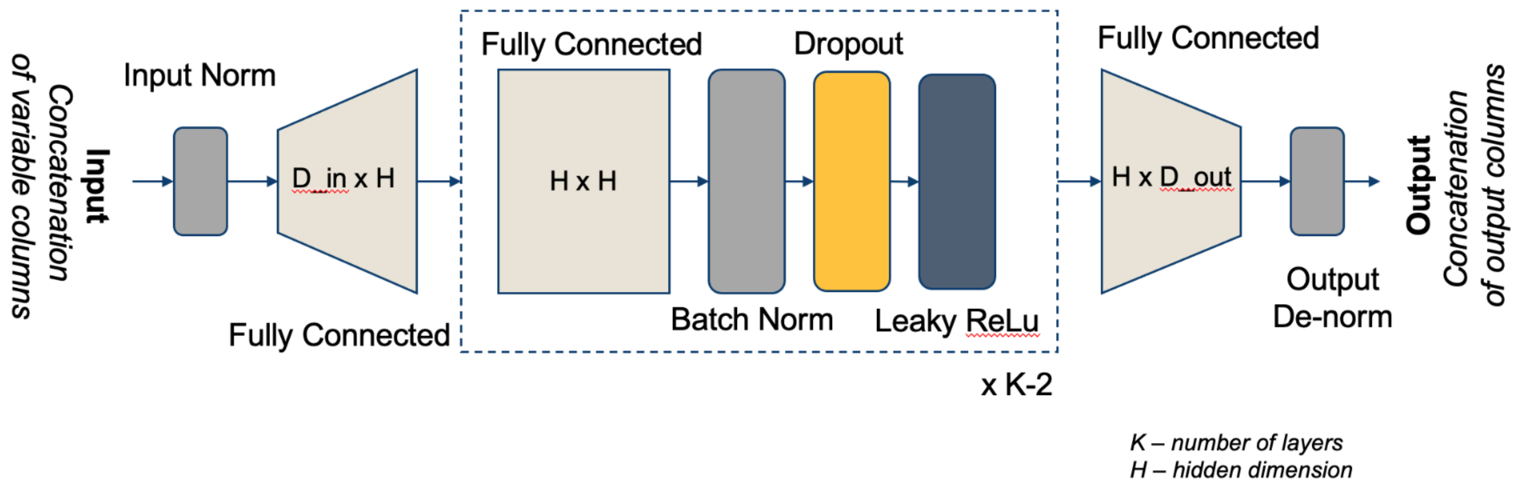
Name	Long Name	shape	unit
PRECT	Total (convective and large-scale) precipitation rate (liq + ice)	(T, 96, 144)	m/s
PRECC	Convective precipitation rate (liq + ice)	(T, 96, 144)	m/s
PTEQ	Q total physics tendency	(T, L, 96, 144)	kg/kg/s
PTTEND	T total physics tendency	(T, L, 96, 144)	K/s

AI Surrogate Architecture

UNSW Climate Change Research Centre

Surrogate Architecture

- 7 FCN (fully connected layers)
- Each layer has a hidden dimension of 512.
- Each layer is followed by batch normalization and dropout with rate of .1



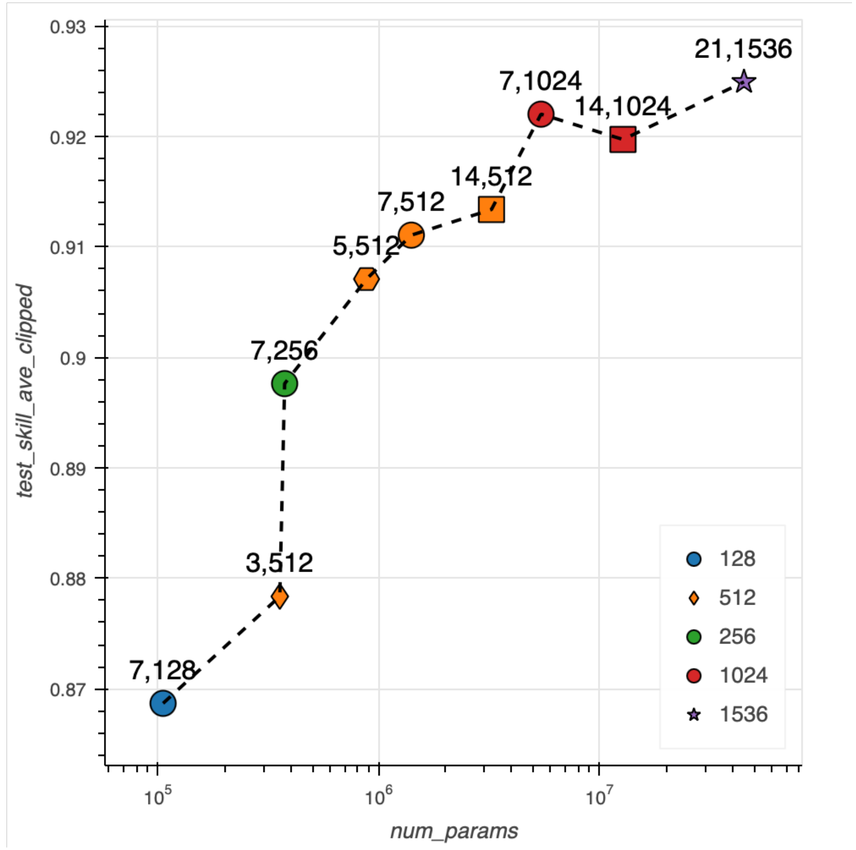
AI Surrogate Model

UNSW Climate Change Research Centre

AI Hyperparameters

We evaluate several parameter values:

- num_layers: 3,5,7,14
- hidden_size: 128, 256, 512, 1024, 1536



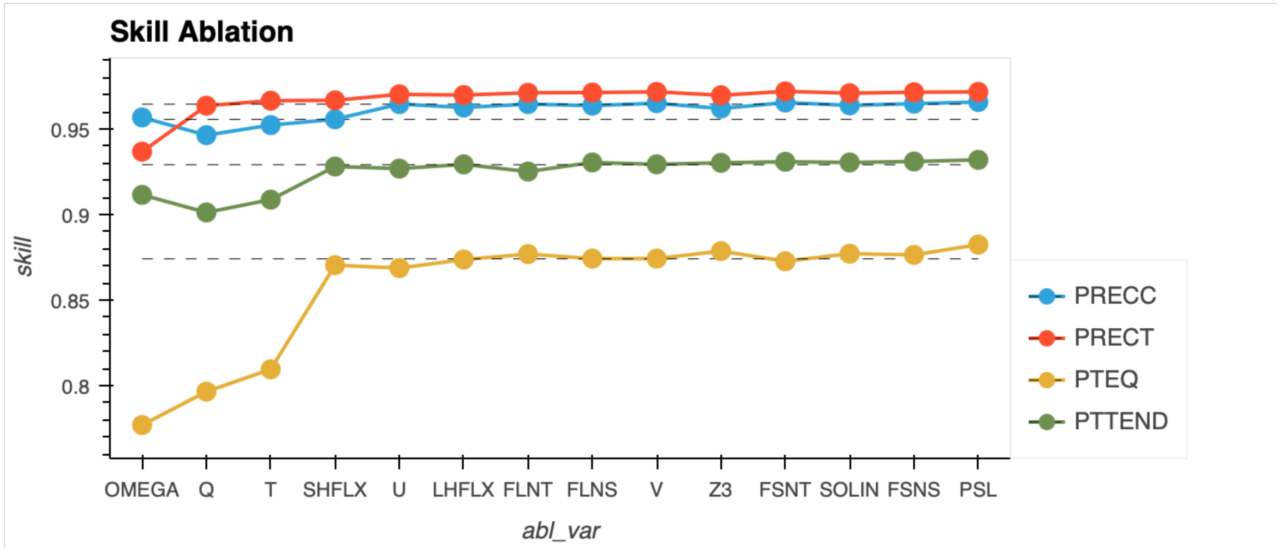
AI Surrogate Ablation study

UNSW Climate Change Research Centre

Input Variable Ablation

Ablation of input variable groups is performed on the surrogate model trained CAM4 data. We measure over skill on a test set as we remove one variable (group) at a time

- Each line corresponds to skill per output variable (note that PTEQ and PTTEND have 26 levels each while PRECC, PRECT are scalars)
- Dashed lines correspond to skill for each output with all input variables
- We observe that OMEGA, Q, T have the greatest impact

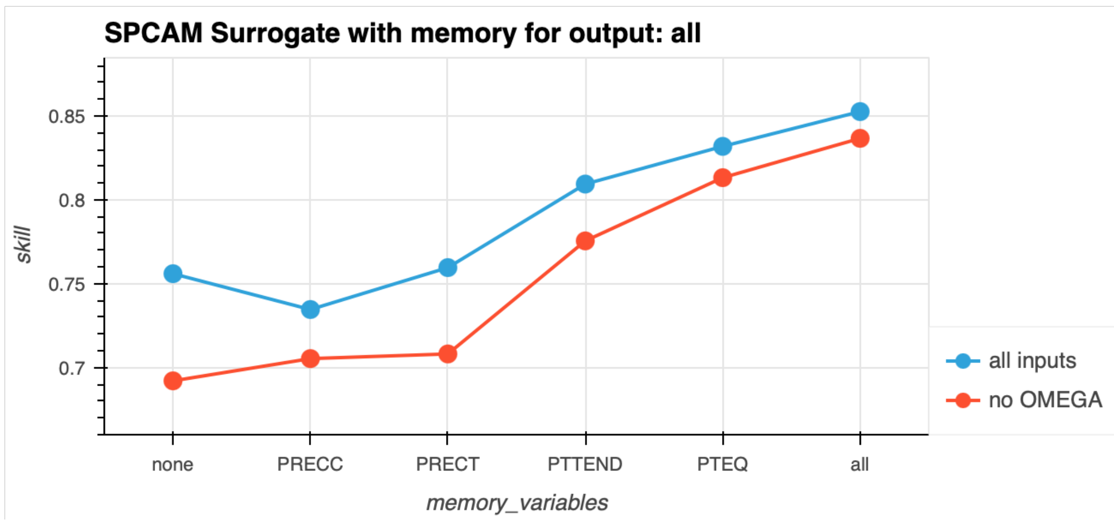


AI Surrogate Model with Memory

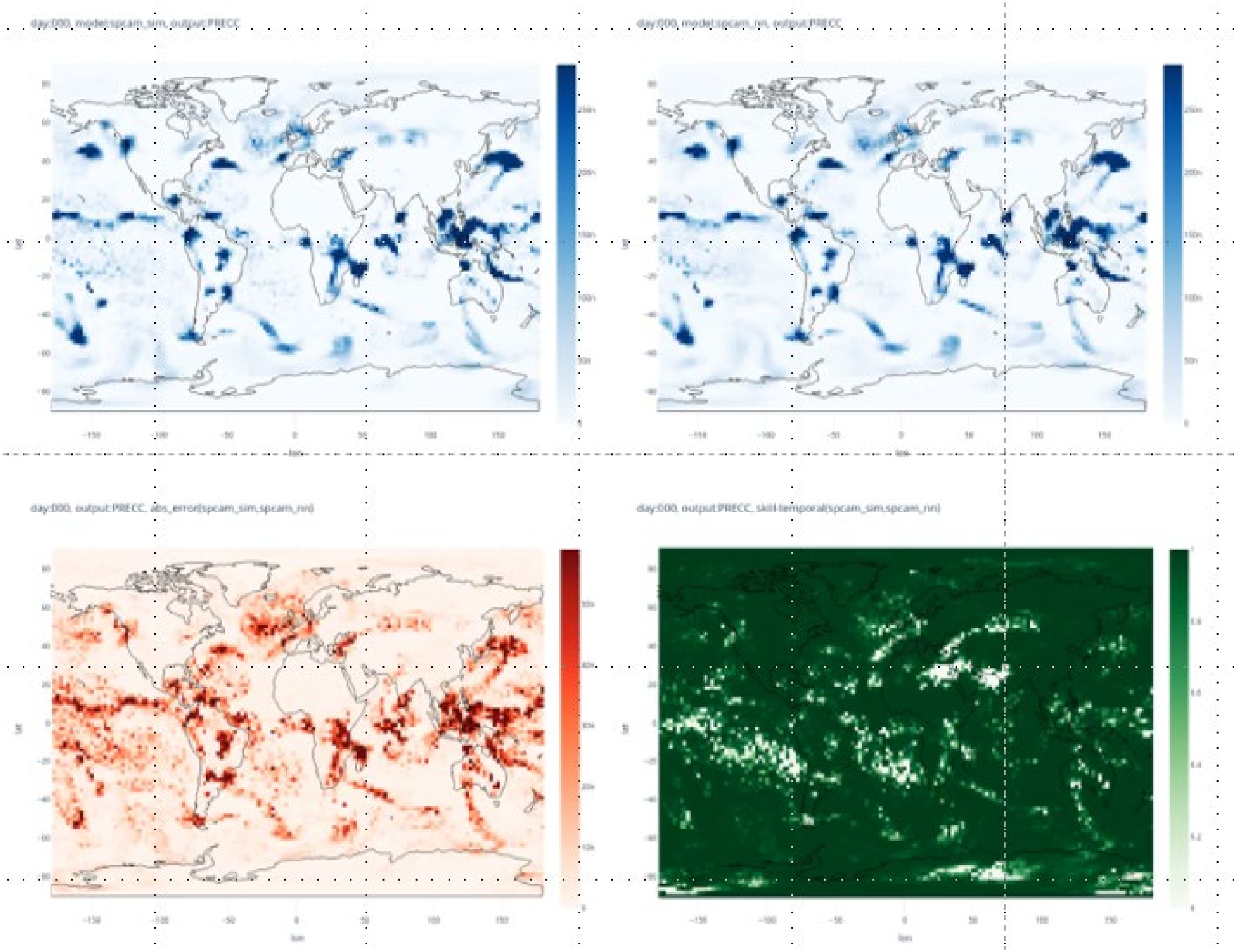
UNSW Climate Change Research Centre

Adding Memory to Surrogate Model

- We test adding output variable from previous time step as an additional input
- We measure improvement in skill for multiple cases (on the SPCAM trained model)
  - Adding one memory variable at a time
  - Adding all variables
  - Same as above but removing OMEGA as an input
- Note PRECT, PRECC are scalars while PTEQ and PTTEND are vector values at 30 levels
- As expected, adding memory of output variable X, improves skill in predicting that variable



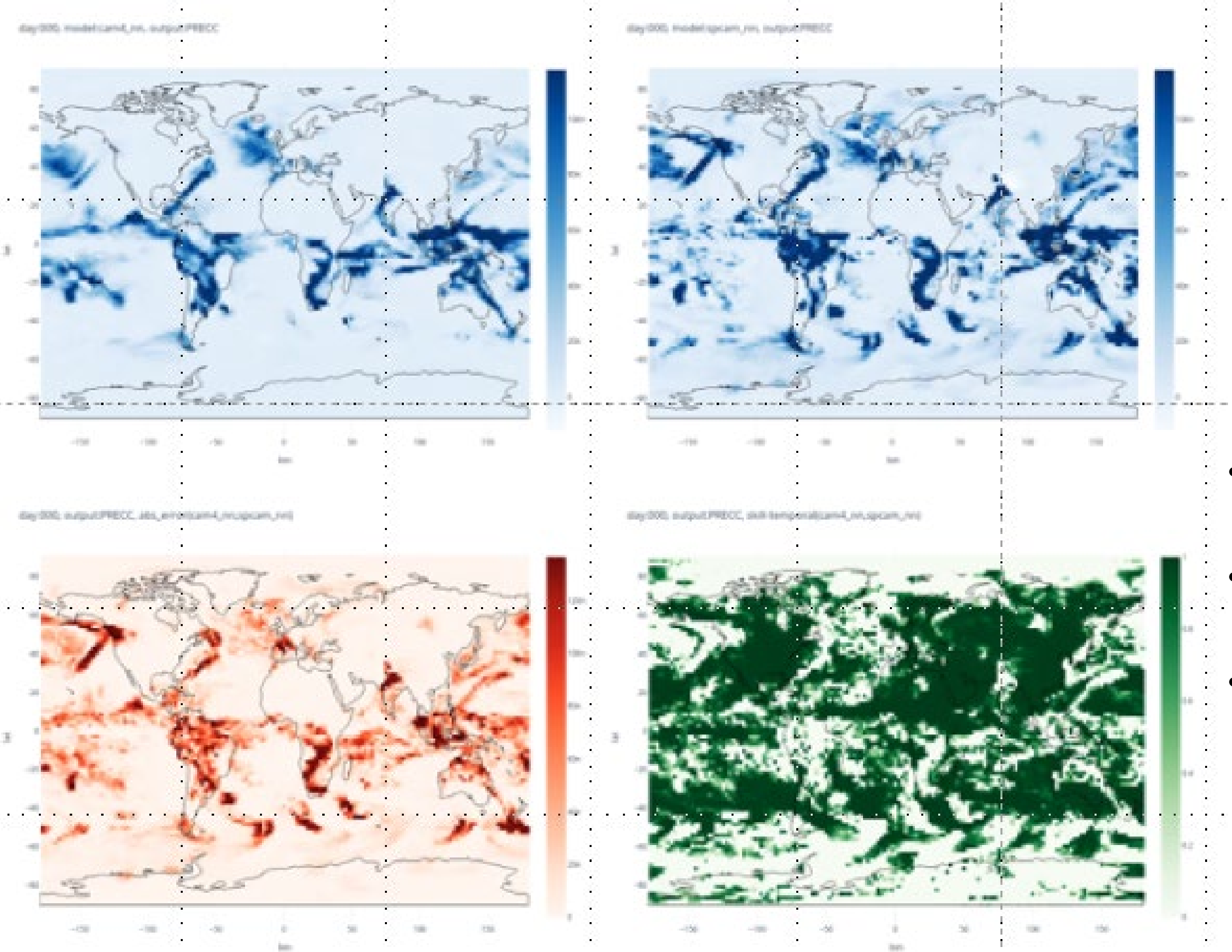
AI example, SPCAM sim vis AI surrogate; PRECC



Videos available on Gaia github

AI example, CAM4- vs SPCAM-trained AI surrogate

UNSW Climate Change Research Centre



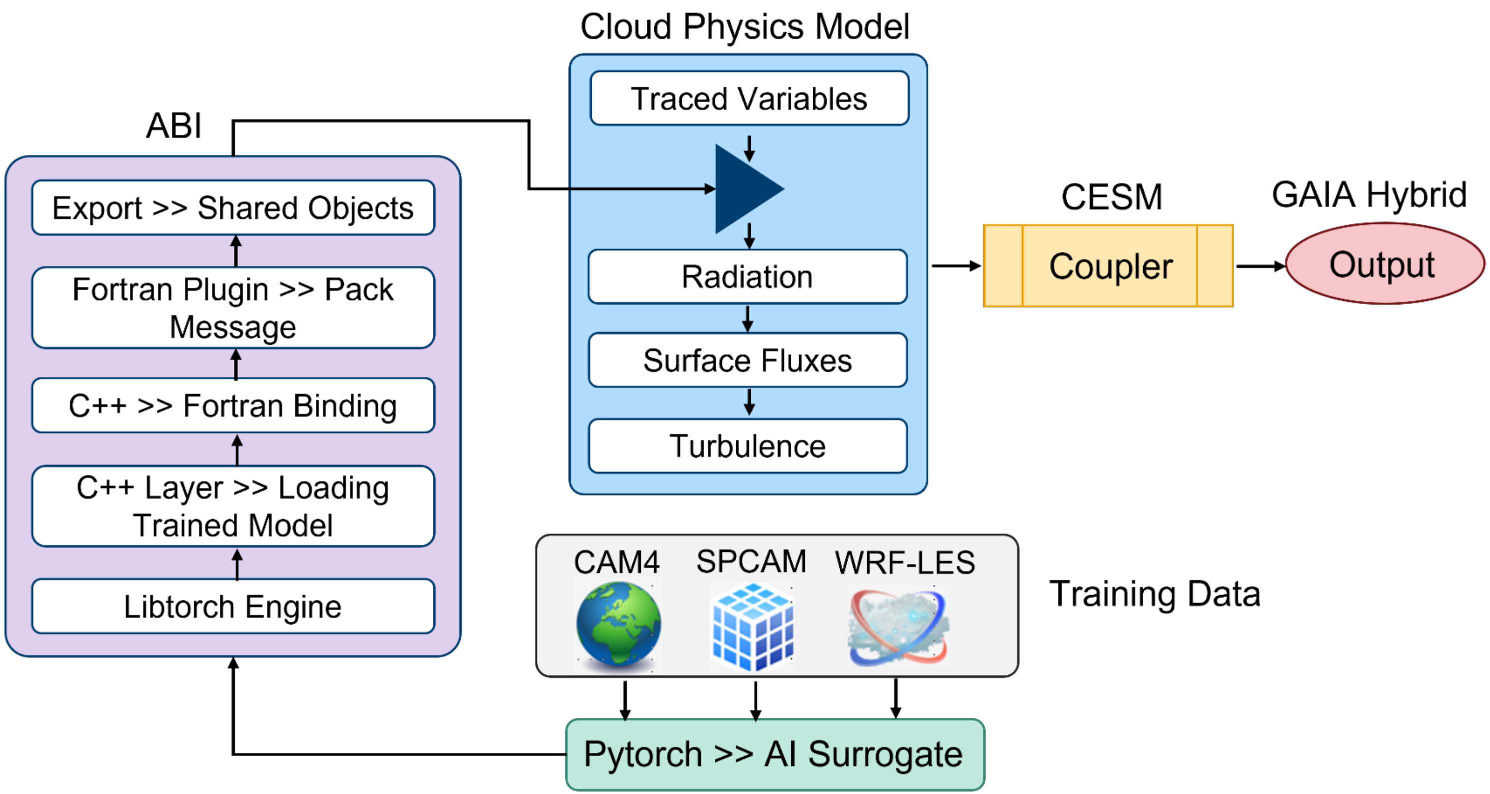
- Overall emulation similarity of ML surrogates to parent GCM is >0.9;
- Similarity skill between surrogates trained on different GCMs is ~0.6-0.7.
- —> Surrogates accurately capture distinctions in GCM physics behavior

Gaia Hybrid Model Architecture

UNSW Climate Change Research Centre

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



Reduced order model/tipping point approach

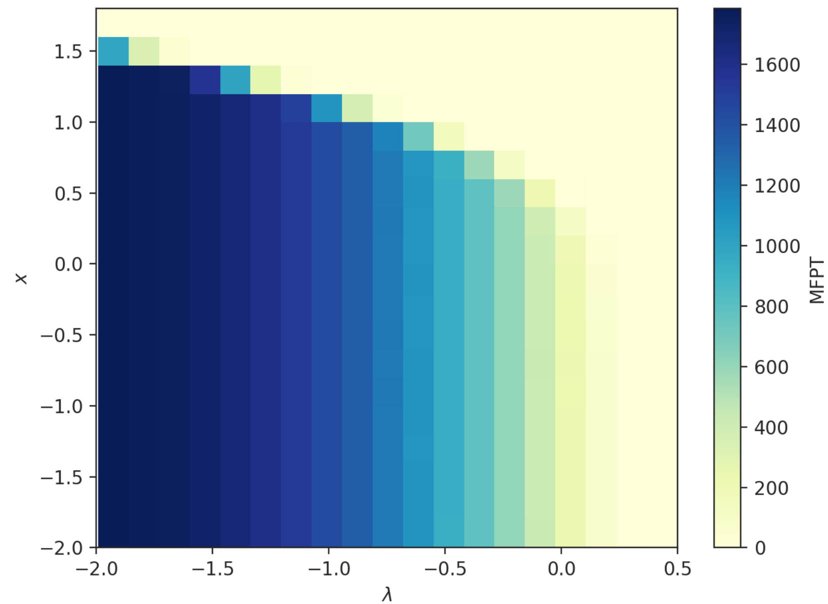
UNSW Climate Change Research Centre

"Technology" READY FOR:

- Learning reduced SDEs in the neighborhood of
  - *turning point* type tipping points
  - *subcritical Hopf* type tipping points
- Computing escape type distributions in the neighborhood of them
- And even attempting to delay/mitigate transition

Something we believe is new:
escape times in phase X parameter space:

For the turning point $dx/dt = \lambda - x^2$
(imagine the parabola passing from (0,0))
top branch unstable / bottom stable)
phase tipping boundary at x=1.5



Achievements to date

str

UNSW Climate Change Research Centre

1. Generated 3xGCM training data; GCM testing and selection
2. Developed and optimised ML emulators of two GCMs
3. Confirmed that emulation skill is very good (compared to precision of existing GCM physics).
4. Established feature importance, including some role for physics memory (still testing)
5. Progress in incorporating PyTorch surrogate into CAM to yield hybrid model (big job)

Challenges and next steps

1. Complete hybrid model engineering
2. Modify hybrid model as needed to obtain stable behavior
3. Use clustering approach to select maximally informative subset of O(100) training cases for LES
4. Run WRF LES on these training cases using CAM boundary conditions
5. Refine retraining strategies for using the LES high-quality training sample to improve surrogate, and execute
6. Retrain (or recalibrate) again on observational data sample (ERA5 and satellite OLR/cloud)
7. Build low dimensional tipping point model based on perturbed (e.g. elevated temp) hybrid model results