FedRAIN-Lite: Federated Reinforcement Algorithms for Improving Idealised Numerical Weather and Climate Models

Pritthijit Nath

Sebastian Schemm
Henry Moss
Peter Haynes
Emily Shuckburgh
Mark Webb (Met Office)

UNIVERSITY OF CAMBRIDGE

NEURAL INFORMATION PROCESSING SYSTEMS

AGU25
New Orleans, LA | 15–19 December 2025

Galveston hurricane (Aug 16, 1915) circulation map (US Weather Bureau) | Climate stripes by Ed Hawkins (Reading)

Overview

I. **Intro to Reinforcement Learning**
   a. Brief-recap of parametrisations
   b. Connecting Go with Climate Model Parametrisations
   c. Why reinforcement learning?
   d. Summary of single-agent RL (Nath et al., 2024)

II. **Multi-Agent RL**
   a. Federated Learning
   b. climateRL – Budyko-Sellers EBM
   c. FedRL Environments – `ebm-v2` and `ebm-v3`
   d. Experiment Pipeline
   e. Results – multi-agent RL
   f. Summary

III. **Next Steps**

Numerical Weather and Climate Modelling

Basics

Operational Weather and Climate Models (such as the UK Met Office's Unified Model) are comprised of **2 MAIN** components:

1. **DYNAMICS** – Fluid motion (solving governing equations) that can be resolved on model's grid.
2. **PHYSICS*** –
   a. Effects of fluid motion smaller than can be resolved (turbulence, convection)
   b. Non-fluid motion (radiation, cloud physics)

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} + 2\boldsymbol{\Omega}\times\boldsymbol{u} + \nabla\Phi + \frac{c_p\theta}{1+\Sigma m_x}\nabla\Pi = \boldsymbol{S_u}$$

Momentum Equation

$$\frac{\partial\theta}{\partial t} + \boldsymbol{u}\cdot\nabla\theta = S_\theta$$

Thermodynamic Equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{u}) = 0$$

Continuity Equation

$$\frac{\partial m_x\rho}{\partial t} + \nabla\cdot(m_x\rho\boldsymbol{u}) = S_x$$

Moisture Transport Equation

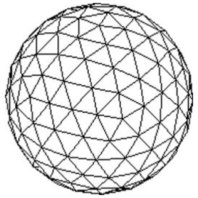
$$\Pi^{\frac{1-\kappa}{\kappa}} + \frac{R}{p_0}\rho\theta = 0$$

Equation of State

Governing Equations

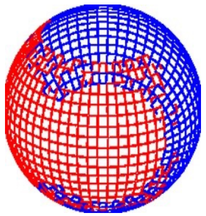
Subgrid Forcings → *Needs to be PARAMETRISED!

Lat-Long Grid
e.g., Unified Model (Met Office), NOAA, NCAR CESM



Icosahedral Grid
e.g., ICON (DWD)
MPAS, NICAM

Cubed Sphere Grid
e.g., LFRic (Met Office)
GFDL FV3



Yin-Yang Grid
e.g., Environment Canada

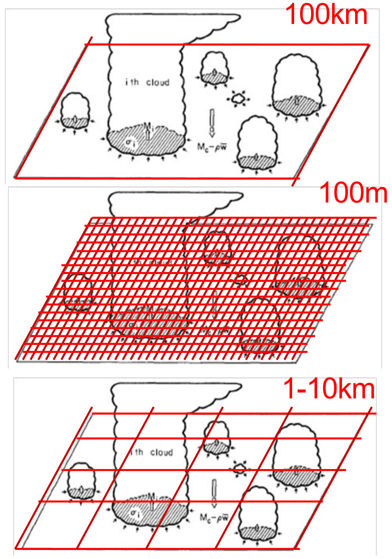
All numerical models are based on a filter, $\theta \rightarrow \bar{\theta}$

$$\frac{\partial\bar{\theta}}{\partial t} + \bar{\boldsymbol{u}}.\nabla\bar{\theta} = -\frac{1}{\rho}\nabla.\overline{\rho\boldsymbol{u}'\theta'} + \frac{L\Pi}{c_p}(\bar{c}-\bar{e}) + Q_R$$

Thermodynamic Equation w/ Filter

Convection parameterisation needs to

1. Represent the effects of sub-filter-scale convection on the filtered flow (first term on RHS).
2. Produce phase-change such as precipitation (second term).
3. Factor in radiative contribution (third term).

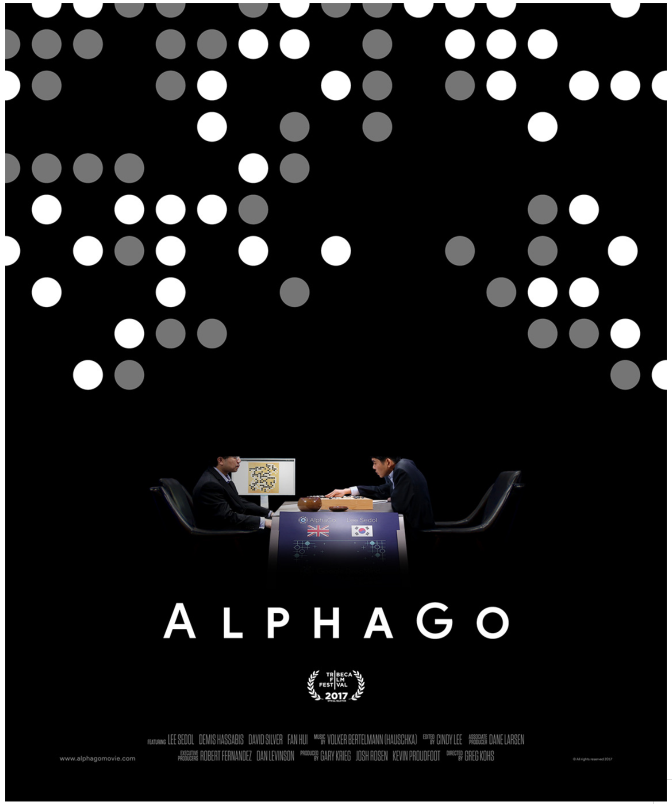


UNIVERSITY OF CAMBRIDGE

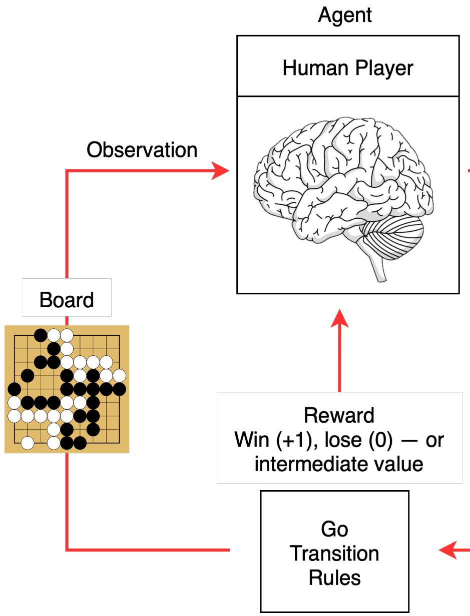
Content adapted from James Kent (Met Office) and Bob Plant (Reading) at the NCAS Climate Modelling Summer School 7 – 19 Sept 2025 in Cambridge, UK

State-Dependent Parametrisations

Go ↔ Climate Modelling



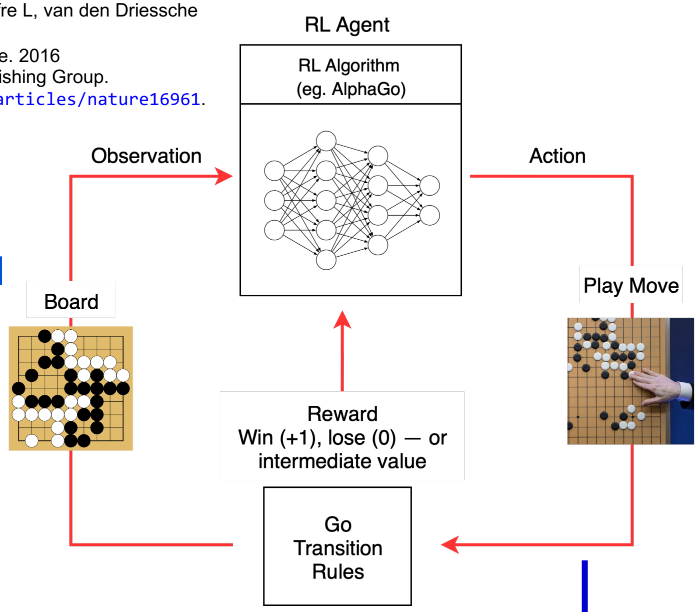
AlphaGo (2017).
https://youtu.be/WXuK6gekU1Y?si=0_Dqe7IGuUz4-dCF

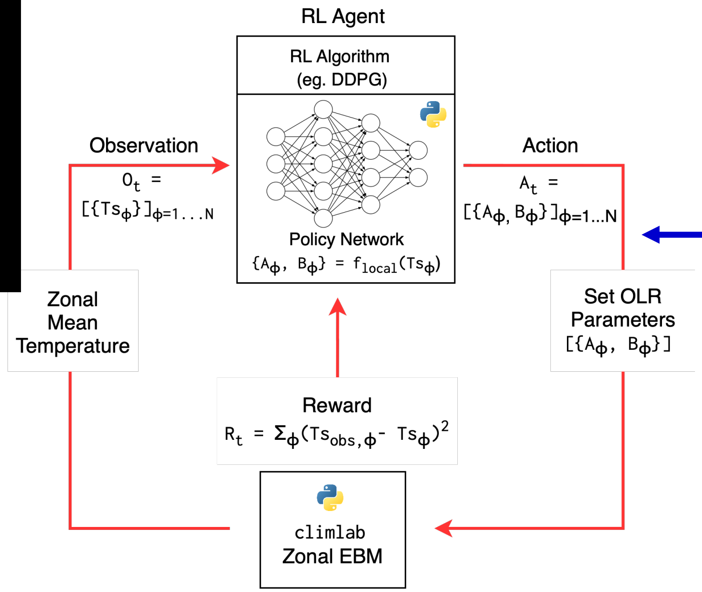


Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature. 2016 Jan;529(7587):484-9. Publisher: Nature Publishing Group. Available from: https://www.nature.com/articles/nature16961.

DeepMind

David Silver





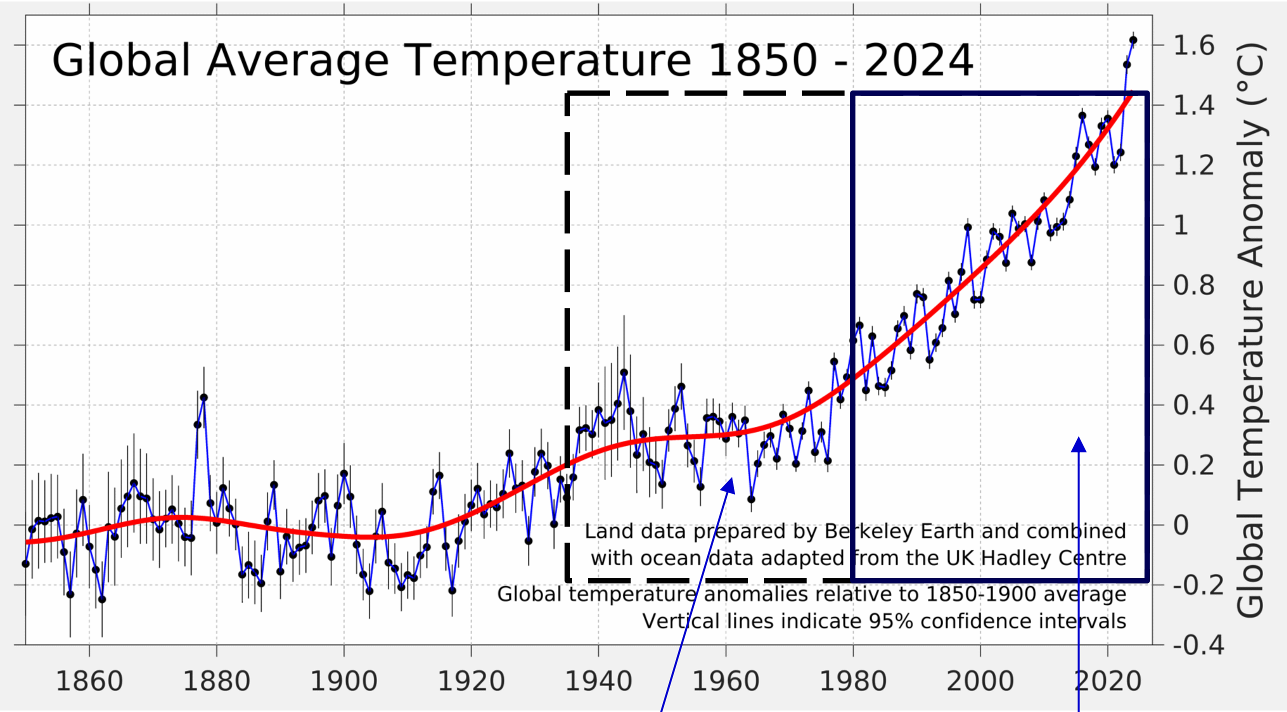
$$C(\phi)\frac{\partial T_s}{\partial t} = (1-\alpha(\phi))Q(\phi) - (A+BT_s) + \frac{D}{\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial T_s}{\partial\phi}\right)$$

Component	Go	Model Parametrisations
State	Board configuration	Climate fields (e.g., Temperature)
Policy	Learnt strategy	Neural network (learnt from experience)
Action	Play move (e.g. Move a stone)	Modify subgrid parameter / flux
Environment	Go transition rules	Climate model (GCM)
Reward	Win (+1) or loss (0) or intermediate value	Skill Score (e.g., RMSE, Bias)

UNIVERSITY OF CAMBRIDGE

Why Reinforcement Learning?

Case AGAINST offline-ML*

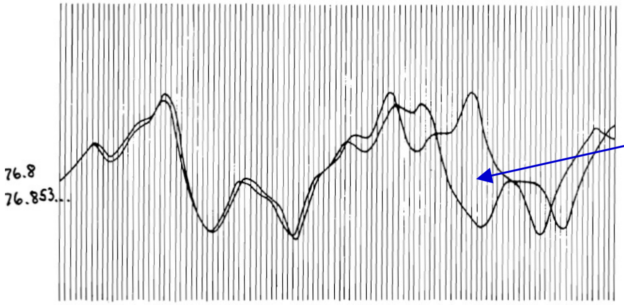


* **ASSUMING** standard off-the-shelf architectures.

ERA5 has been extended to 1940. In total – 84 years of hourly global data.

Most ML data is observed from the satellite era – 1980 onwards.

1. Global warming is causing temperature and climate to go **OUT-OF-DISTRIBUTION**. No guarantees for trained distribution to hold for warming regimes. **EXPENSIVE RE-TRAINING** is required for new data.

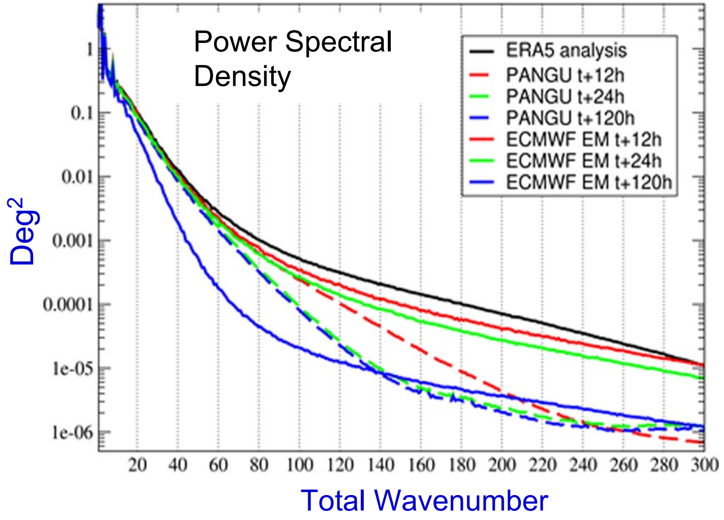


How two weather patterns diverge. From nearly the same starting point, Edward Lorenz saw his computer weather produce patterns that grew farther and farther apart until all resemblance disappeared. (From Lorenz's 1961 printouts.)

2. Atmospheric dynamics is **chaotic** … creating a static dataset can be computationally **VERY EXPENSIVE.**

3. End-to-End Models (e.g., GraphCast, Aurora) **DO NOT EXPLICITY RESOLVE** conservation physics. No guarantees for conservation in **LONG-TERM** Rollouts and often causes **BLURRING**. Need physics informed architectures.

Bonavita M. On some limitations of data-driven weather forecasting models. arXiv; 2023. ArXiv:2309.08473 [stat]. Available from: https://arxiv.org/abs/2309.08473.



4. **LACK OF** adaptability **AGAINST** sparse-rewards (e.g., combining 1-hr ERA5 data with 6-hour satellite data).

5. **NOT OPTIMISED** for long-term forecasting. Most architectures predict one-step-ahead and produce multiple outputs through auto-regressive forecasts.

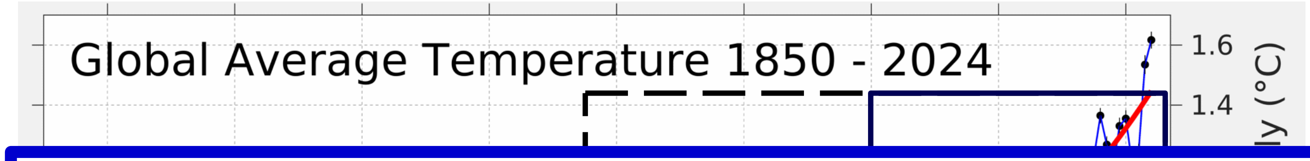
Reinforcement Learning (RL) can:

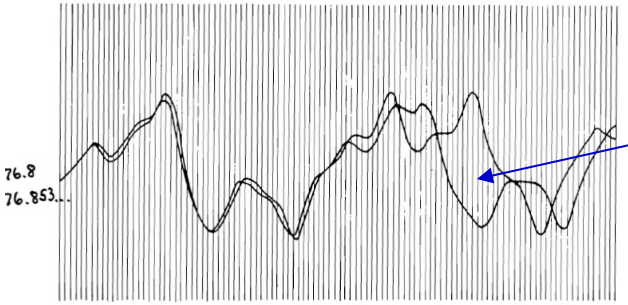
1. **AVOID** the need for costly offline datasets through online learning.
2. **ADAPT** continuously as new data emerge.
3. **LEVERAGE** the underlying physics-based climate model to ensure conservation.
4. **LEARN** effectively even from sparse or delayed rewards.
5. **OPTIMISE** directly for long-term rollouts consistent with climate dynamics.

UNIVERSITY OF CAMBRIDGE

Why Reinforcement Learning?

Case AGAINST offline-ML*





How TWO WEATHER PATTERNS DIVERGE. From nearly the same starting

2. Atmospheric dynamics is **chaotic** … creating a static dataset can be computationally **VERY EXPENSIVE.**

<u>DIFFERENTIABLE</u> ESMs ARE <u>NOT</u> NEEDED

so are GPUs :)

* **ASSUMING** standard off-the-shelf architectures.

ERA5 has been extended to 1940. In total – 84 years of hourly global data.

Most ML data is observed from the satellite era – 1980 onwards.

1. Global warming is causing temperature and climate to go **OUT-OF-DISTRIBUTION**. No guarantees for trained distribution to hold for warming regimes. **EXPENSIVE RE-TRAINING** is required for new data.

4. **LACK OF** adaptability **AGAINST** sparse-rewards (e.g., combining 1-hr ERA5 data with 6-hour satellite data).

5. **NOT OPTIMISED** for long-term forecasting. Most architectures predict one-step-ahead and produce multiple outputs through auto-regressive forecasts.

Reinforcement Learning (RL) can:
1. **AVOID** the need for costly offline datasets through online learning.
2. **ADAPT** continuously as new data emerge.
3. **LEVERAGE** the underlying physics-based climate model to ensure conservation.
4. **LEARN** effectively even from sparse or delayed rewards.
5. **OPTIMISE** directly for long-term rollouts consistent with climate dynamics.

UNIVERSITY OF CAMBRIDGE

Summary

RAIN (Nath et al., 2025) : Parametrisations using Single-agent RL

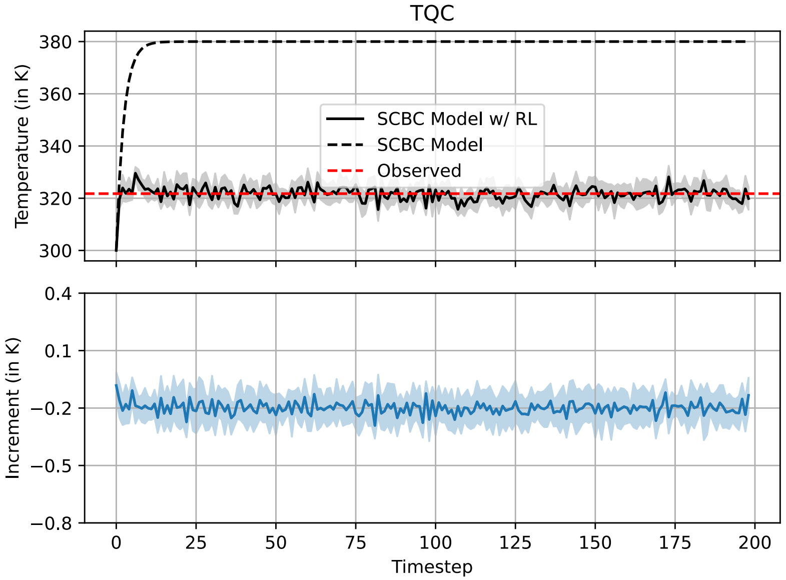
1. **EVALUATED** model-free RL across two idealised climate setups, adjusting parameters via policies informed by evolving states.

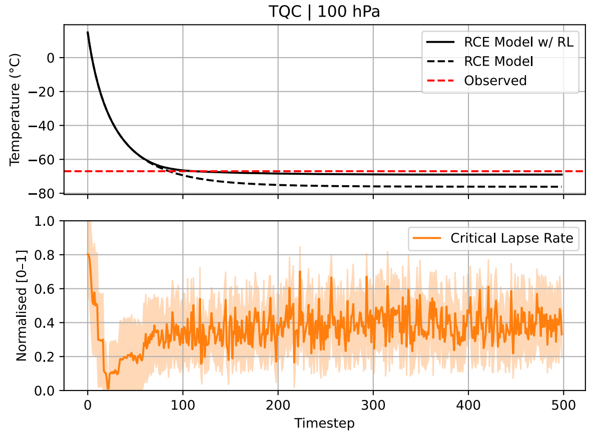
2. **OBSERVED** mostly off-policy methods (DDPG, TD3, TQC) to perform best across environments (except DPG in RCE).

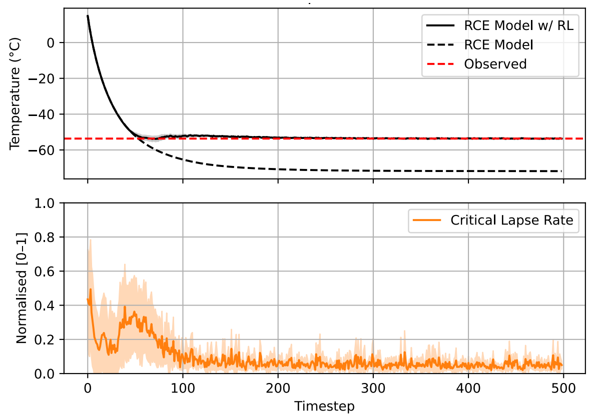
3. **HIGHLIGHT** understanding dynamics is key under computational constraints of climate simulations.

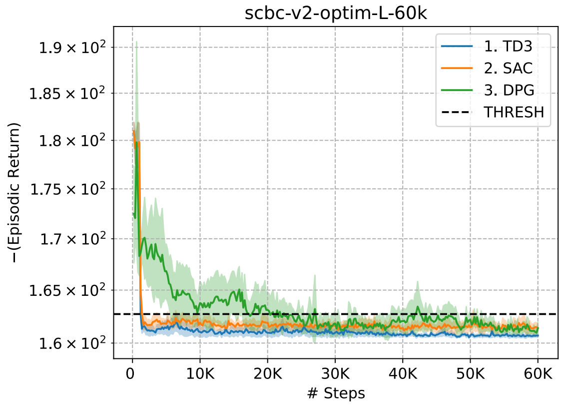
4. **DEVELOPED** a reproducible experimental workflow as a foundation for RL-based hybrid parameterisations in complex climate models.

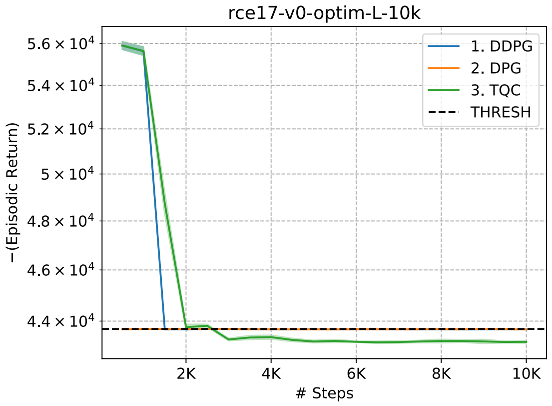
5. **IDENTIFIED** structural limitations of the reward formulation (e.g., mean squared error) can cause over-compensation errors.

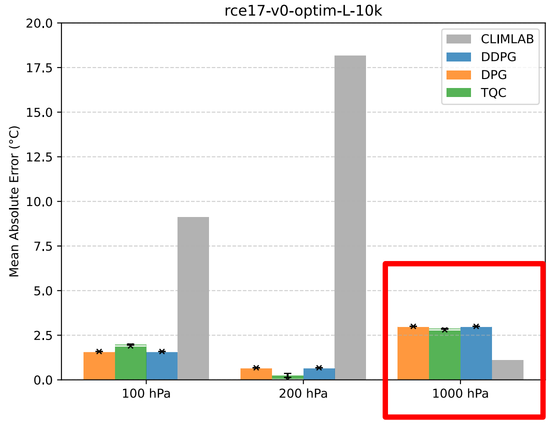


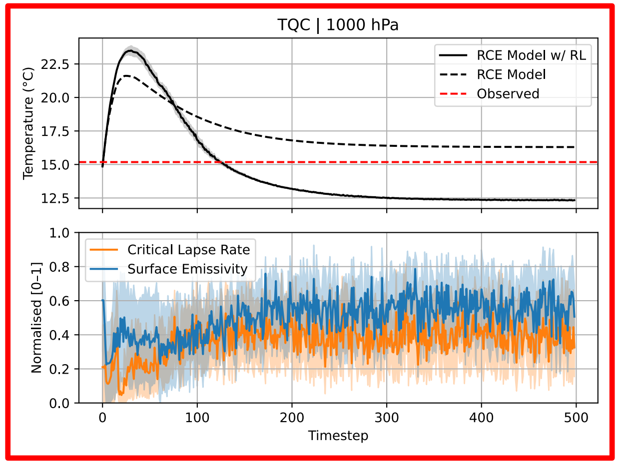












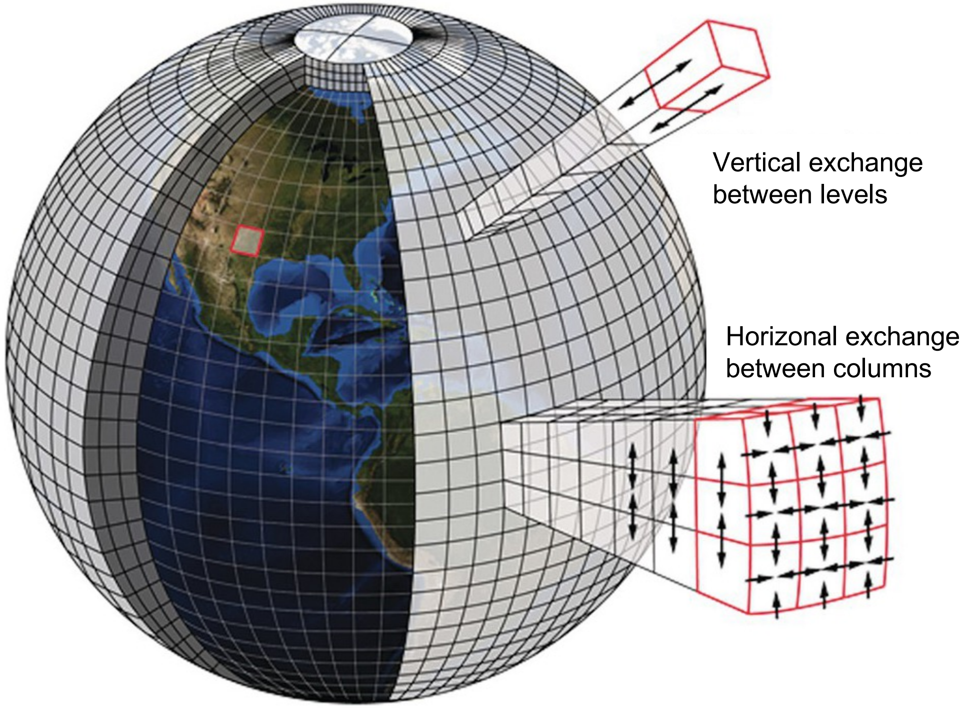
UNIVERSITY OF CAMBRIDGE

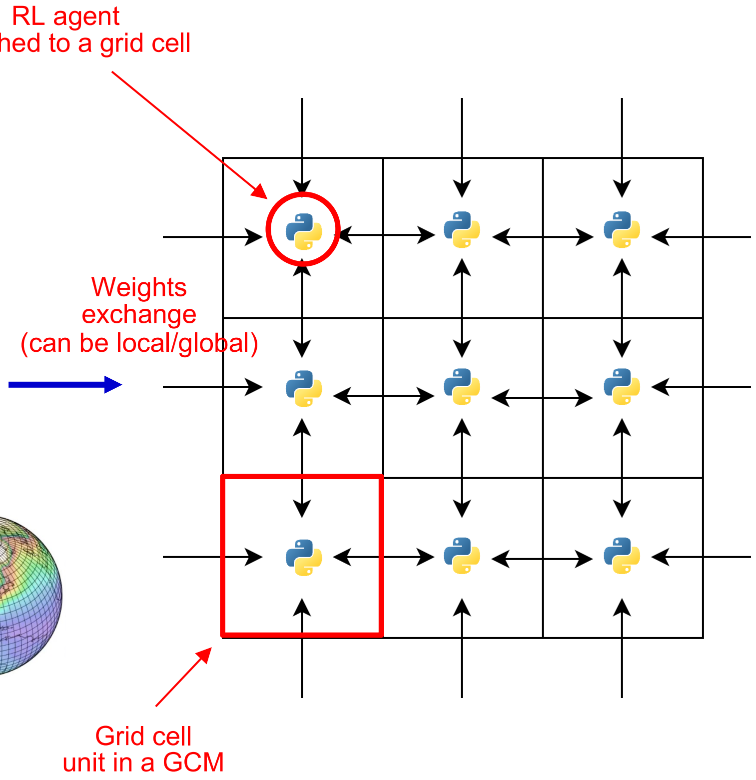
[1] Nath P, Moss H, Shuckburgh E, Webb M. RAIN: Reinforcement algorithms for improving numerical weather and climate models. 2025 EGU General Assembly (Oral); 2025. EGU25-5159 (ITS1.4/CL0.10). Available from: https://arxiv.org/abs/2408.16118.

[2] Nath P, Schemm S, Moss H, Haynes P, Shuckburgh E, Webb M. Making Tunable Parameters State-Dependent in Weather and Climate Models with Reinforcement Learning (Manuscript in preparation).

Spatial Decomposition

Federated Learning





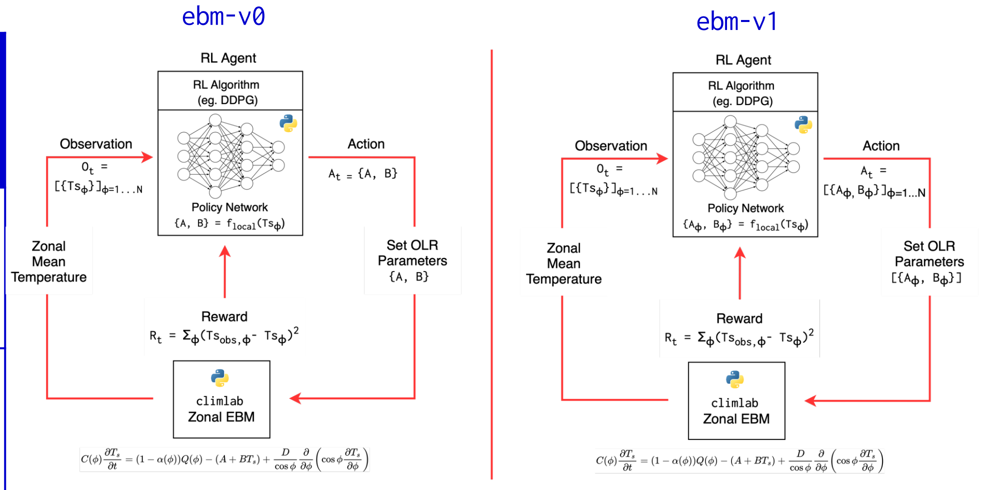
UNIVERSITY OF CAMBRIDGE

Discrete Climate Model Grid Schematic (Earth Magazine) | Astronaut riding a horse (OpenAI DALL-E, MIT Tech Review)

climateRL Environments

Budyko-Sellers Energy Balance Model (EBM)

Component	Budyko-Sellers Energy Balance Model (EBM)
State	Temperature $T_s(\phi)$
Policy	Neural network (learnt from experience) π_θ
Action	Modify albedo $\alpha(\phi)$ α_0 α_2, diffusive transport D, OLR coefficients A B
Environment	EBM model $C(\phi)\frac{\partial T_s}{\partial t} = \underbrace{(1-\alpha(\phi))Q(\phi)}_{\text{absorbed shortwave}} - \underbrace{(A+BT_s)}_{\text{longwave cooling}} + \underbrace{\frac{D}{\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial T_s}{\partial\phi}\right)}_{\text{diffusive transport}}$
Reward	$r(t) = -\frac{1}{N}\sum_{i=1}^{N}\left[T_{\text{simulated}}(t,\phi_i) - T_{\text{observed}}(\phi_i)\right]^2$

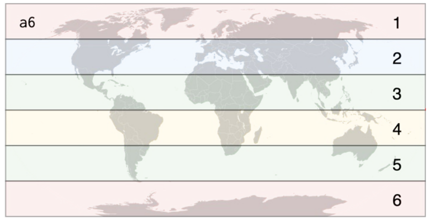


ebm-v0: 5 parameter values same for all 96 latitudes
ebm-v1: 96 A and B pairs 1 – for each latitude
ebm-v2: FedRL version, global input but regional rewards (ensemble)
ebm-v3: ebm-v3 + sliced regional inputs (GCM-like)

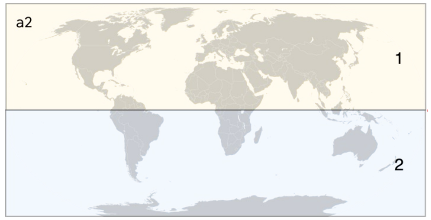
Parameter	Low	High	Canonical
A	140	420	210
B	1.95	2.05	2
α_0	0.3	0.4	0.354
α_2	0.2	0.3	0.25
D	0.55	0.65	0.6

Federated Reinforcement Learning (FedRL)

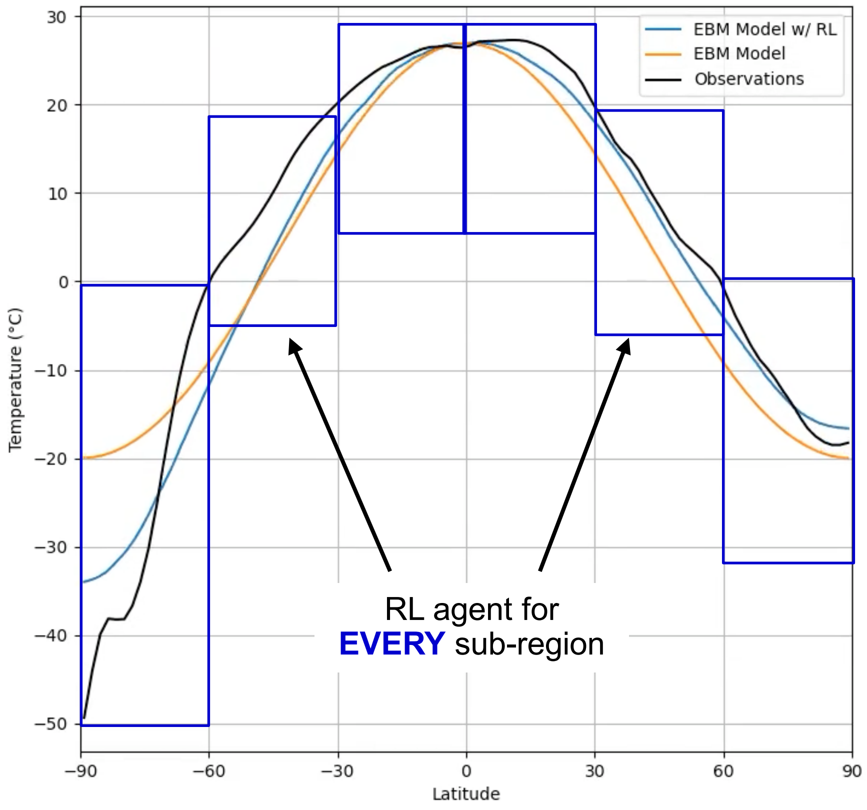
ebm-v2 Environment



6-region decomposition



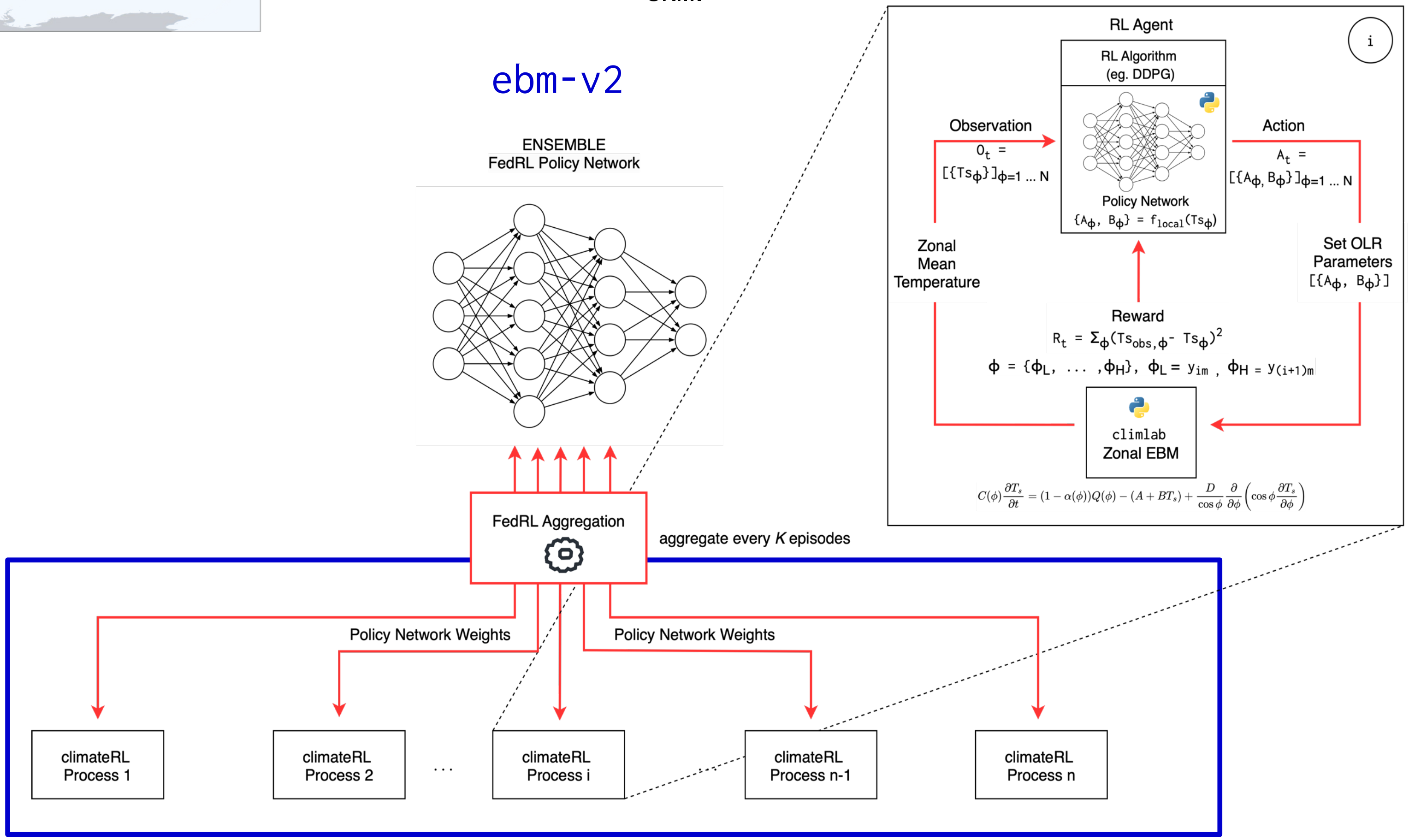
2-region decomposition



Federated RL can provide free improvements in existing RL parameterisations through weight-sharing.

FedRAIN — leveraging the GCM structure, can:

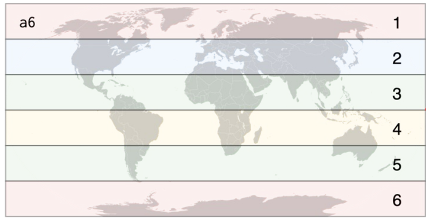
1. **LATCH** an RL agent onto each regional climate process via GCM spatial decomposition.
2. **SHARE** weights periodically across agents for **COLLECTIVE** learning.
3. Perform **FASTER** convergence and more **LOCALISED** skill.



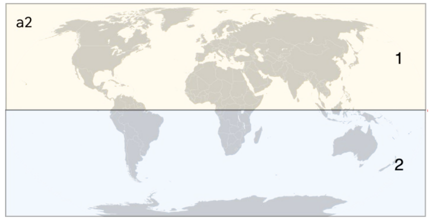
UNIVERSITY OF CAMBRIDGE

Federated Reinforcement Learning (FedRL)

`ebm-v3` Environment



6-region decomposition

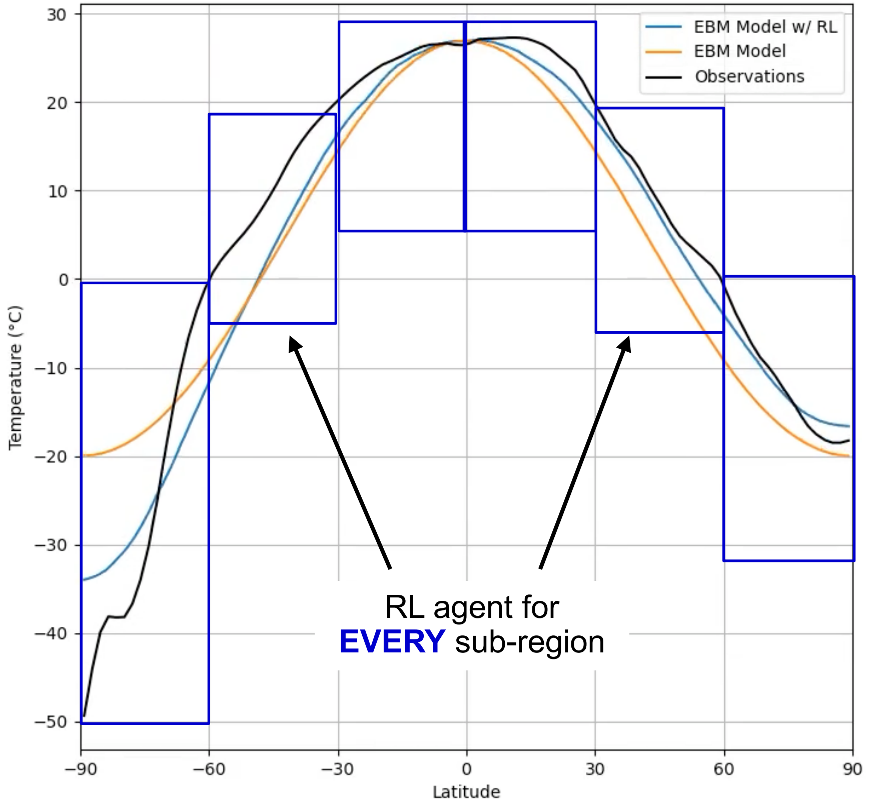


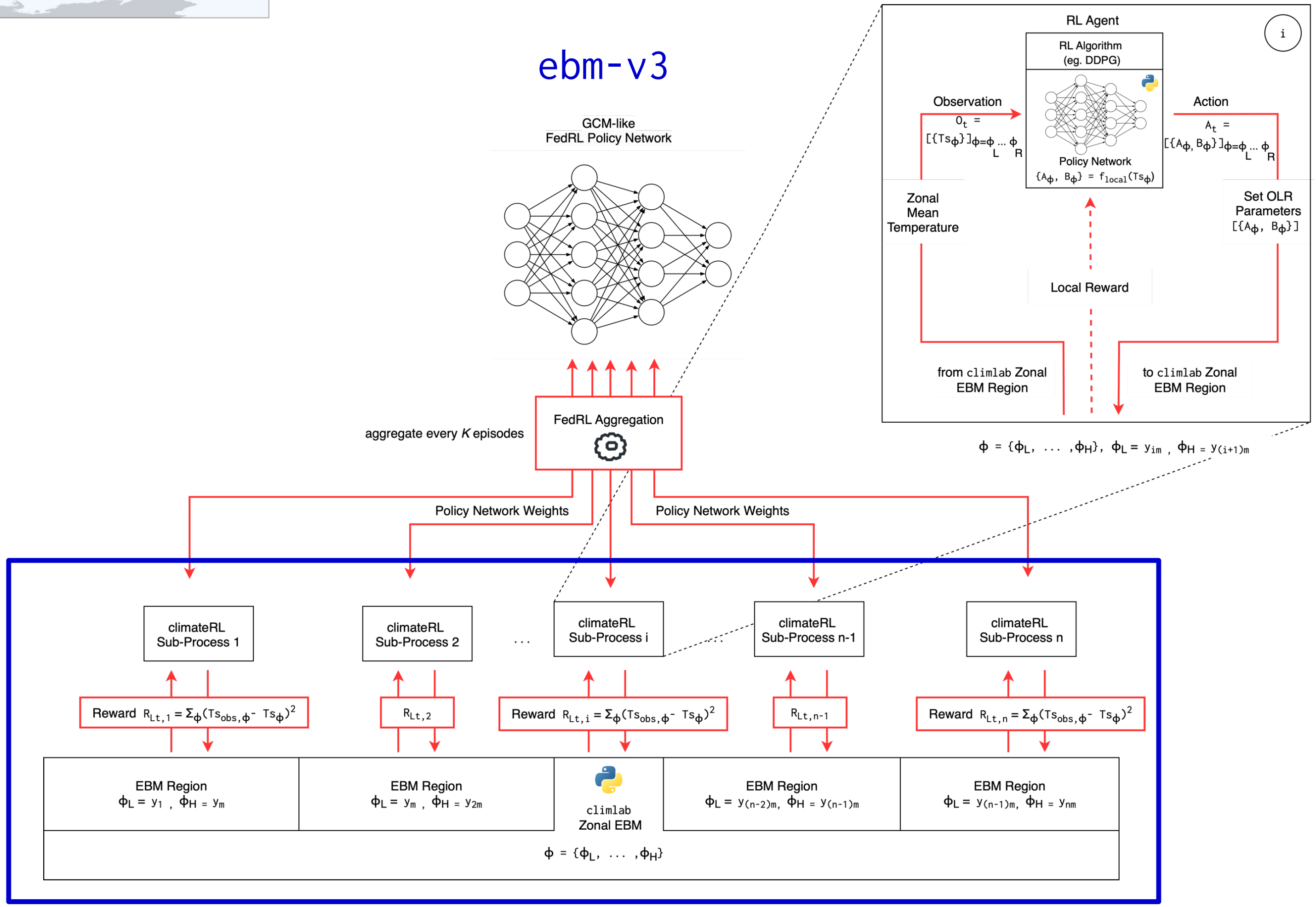
2-region decomposition

Federated RL can provide free improvements in existing RL parameterisations through weight-sharing.

FedRAIN — leveraging the GCM structure, can:

1. **LATCH** an RL agent onto each regional climate process via GCM spatial decomposition.
2. **SHARE** weights periodically across agents for **COLLECTIVE** learning.
3. Perform **FASTER** convergence and more **LOCALISED** skill.

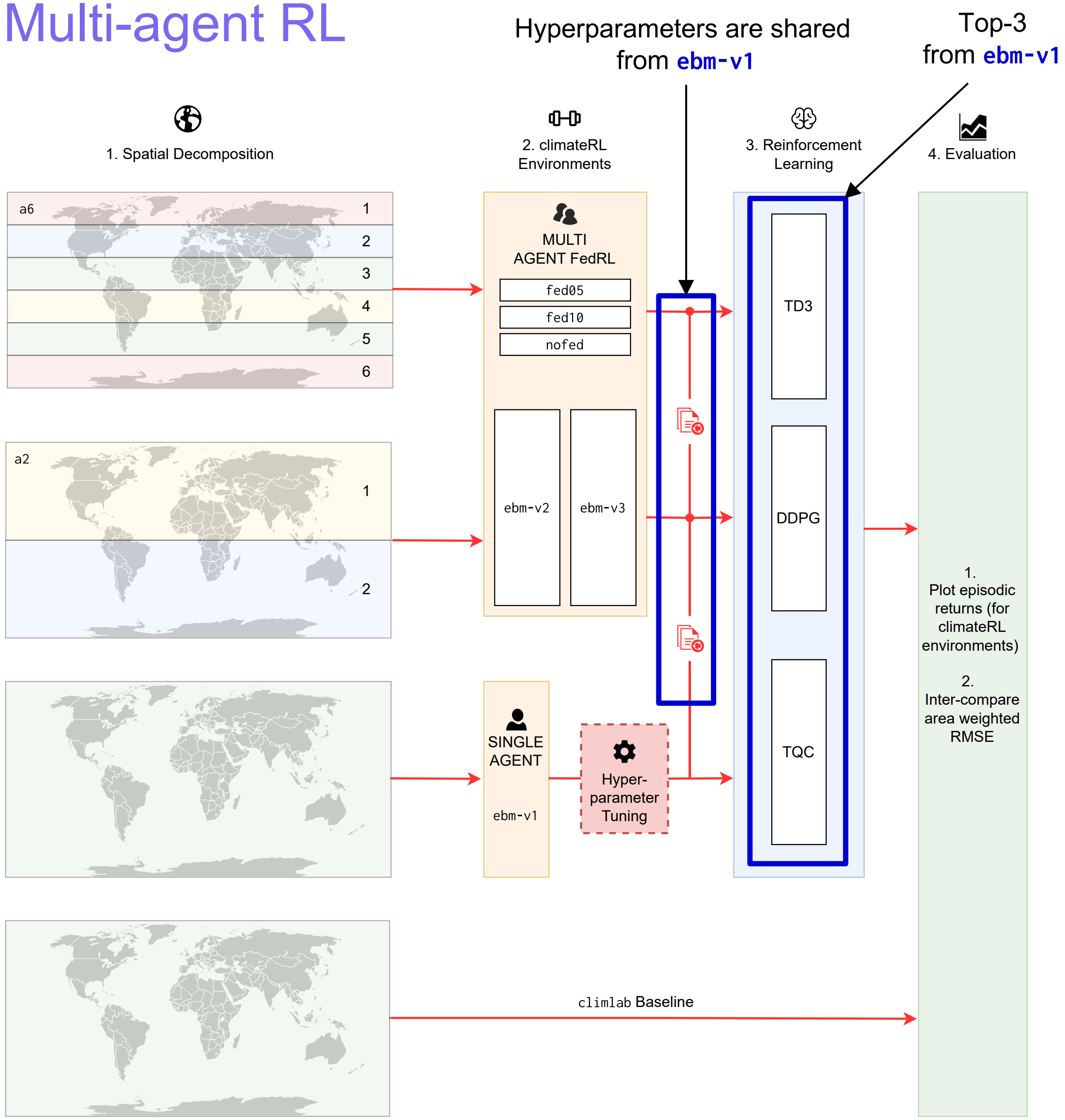




UNIVERSITY OF
CAMBRIDGE

Experiment Pipeline

Multi-agent RL



Code for Multi-agent RL
`https://github.com/p3jitnath/climate-rl-fedrl`

Code for FedRAIN API
`https://github.com/p3jitnath/climate-rl-fedrain-api`

UNIVERSITY OF CAMBRIDGE

Results

Multi-agent EBM Regional Skill

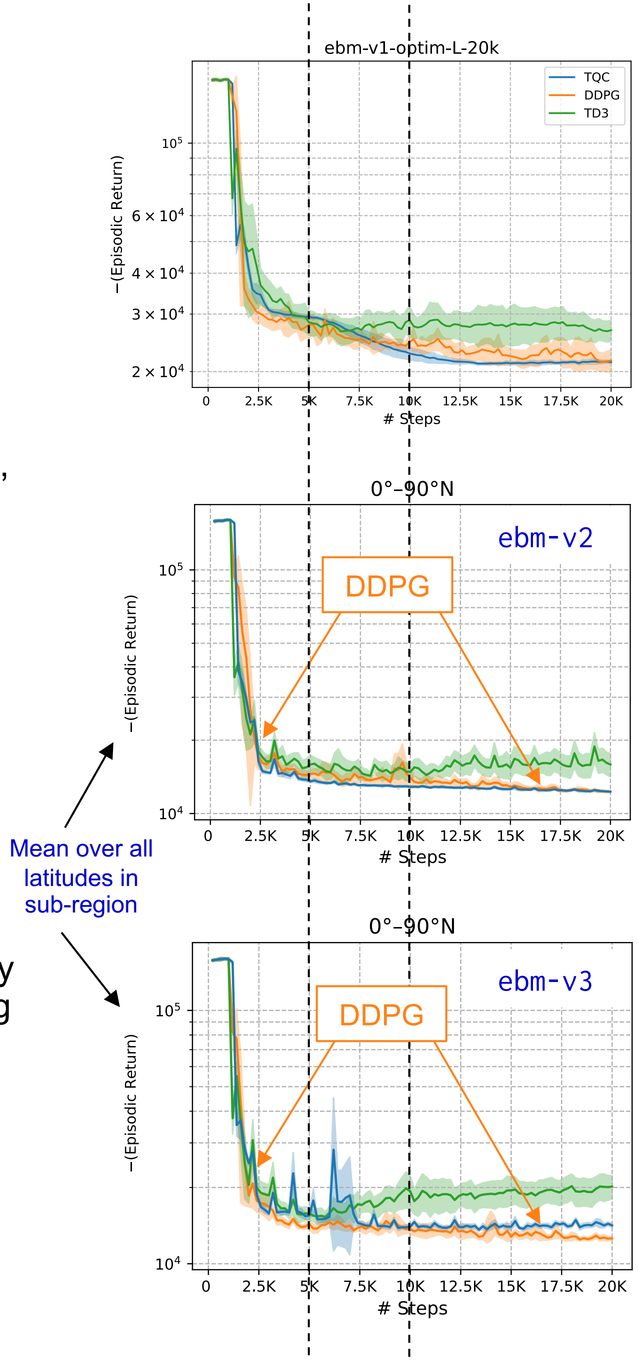
1. `fed05` in `a2` achieves faster (5k vs. 10k) and more stable convergence than the single-agent baseline.

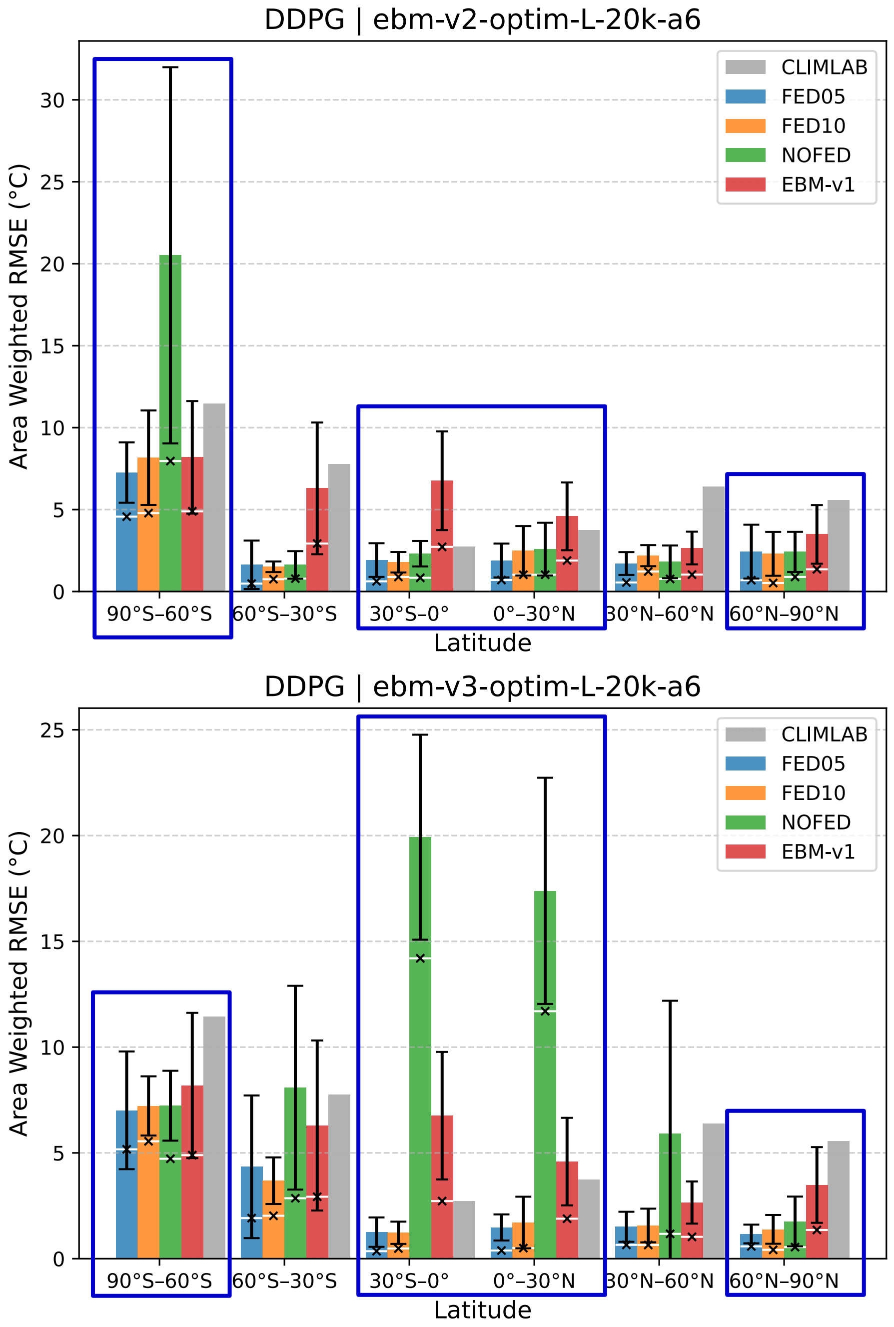
2. `fed05` consistently outperforms both the static baseline `climlab` and unfederated `nofed` setups, showing the strongest gains (vs. `ebm-v1`) in tropical regions (30°S–0° and 0°–30°N).

3. Polar regions show comparable or better skill than the single-agent `ebm-v1`, demonstrating improved handling of sharp gradients through local specialisation.

4. Federated coordination (with DDPG) improves stability and reduces error variance, particularly in `ebm-v3`, where localised learning enhances performance.

5. Fine-grained decompositions increase sensitivity to hyperparameters, with TD3 and TQC showing transient collapses under modified inputs or rewards.





UNIVERSITY OF CAMBRIDGE

Results

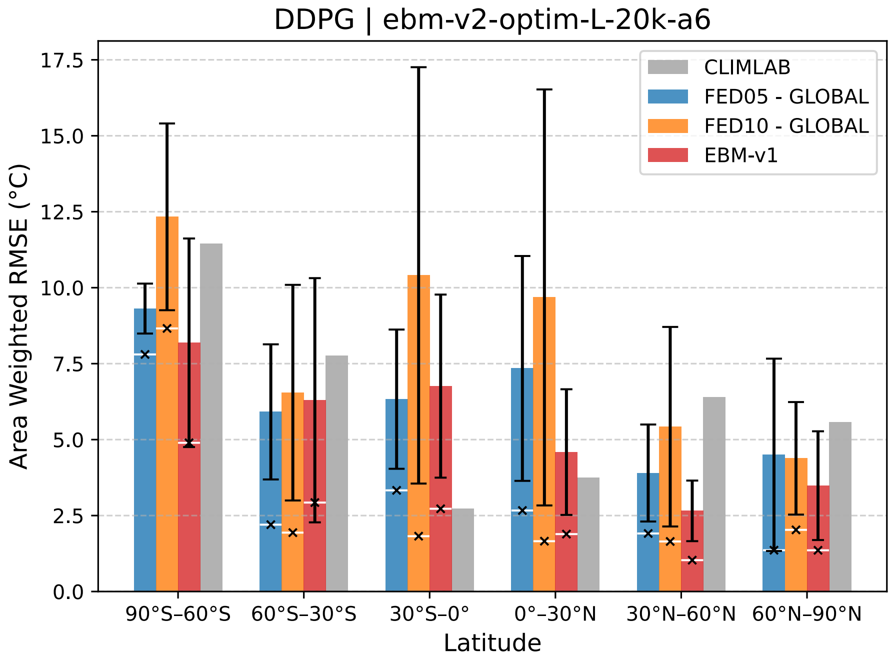
Multi-agent EBM Global Skill

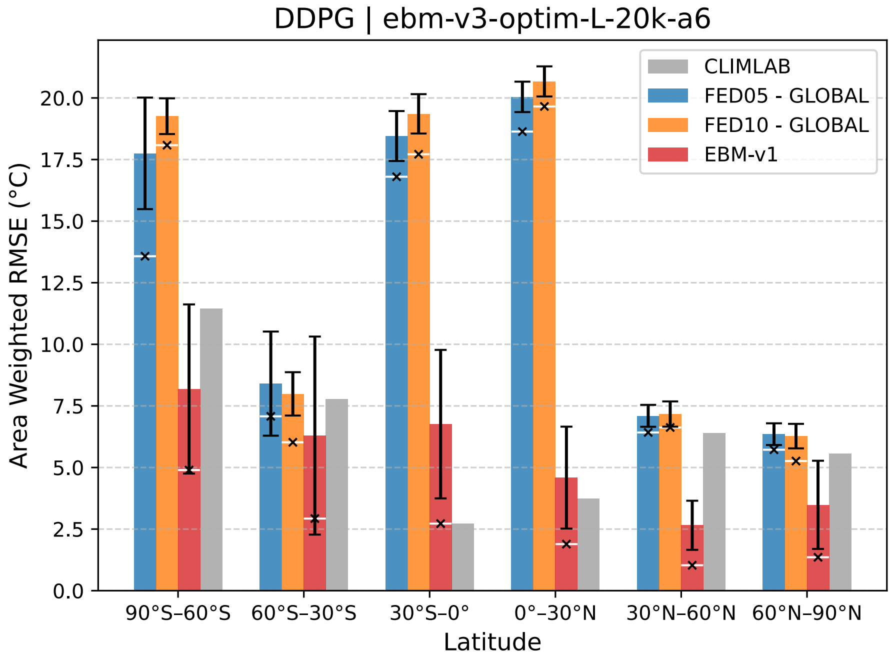
1. Global aggregation of non-local policies reduces performance, increasing both areaWRMSE and inter-seed variance compared to local evaluations.

2. Regional specialisation benefits are lost when policies are merged globally, with `fed10-GLOBAL` performing worst due to infrequent synchronisation.

3. In `ebm-v2`, modest tropical and mid-latitude gains persist (vs. `ebm-v1`), but in `ebm-v3` global inference further degrades robustness below the `climlab` baseline.

4. `fed05` slightly improves stability, highlighting the challenge of reconciling heterogeneous climate regimes under global coordination.



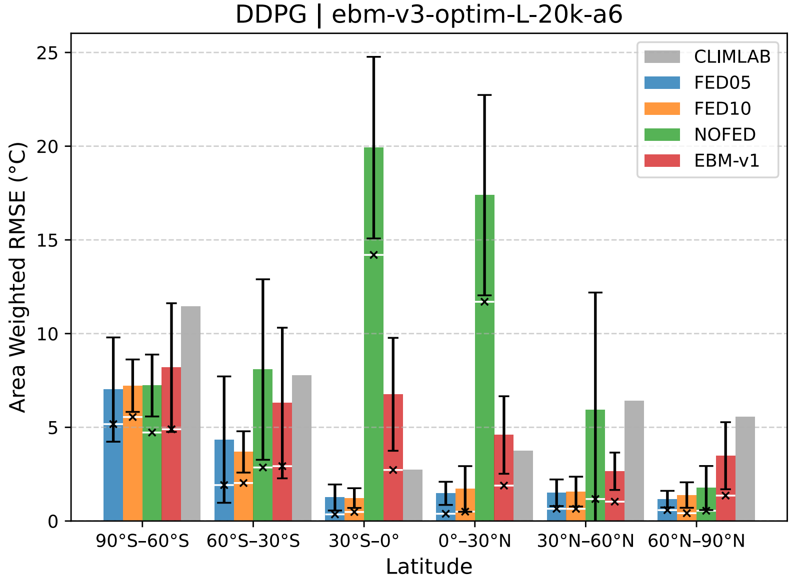


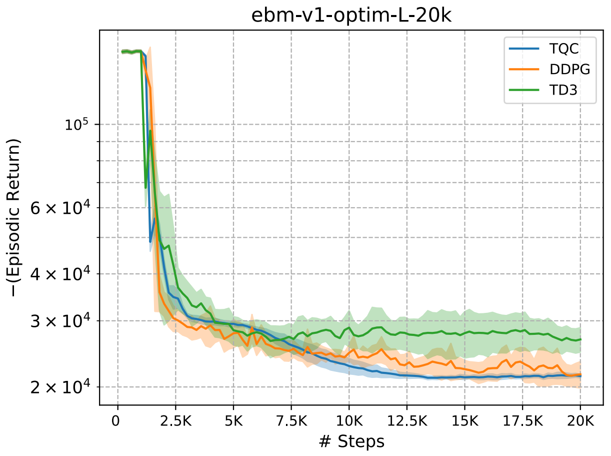
UNIVERSITY OF CAMBRIDGE

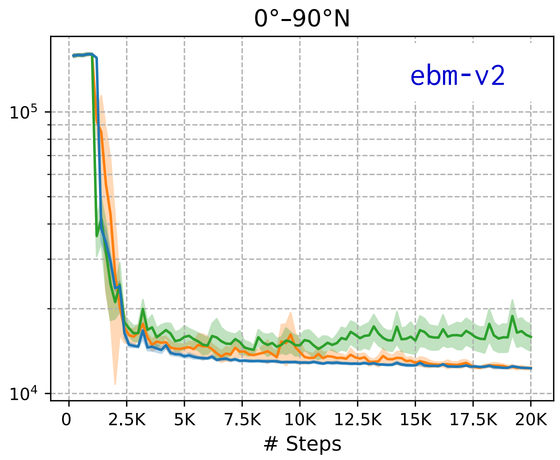
Summary

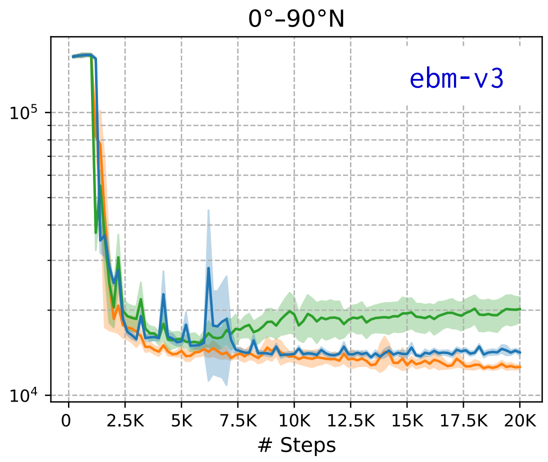
FedRAIN-Lite (Nath et al., 2025) : Parametrisations using FedRL

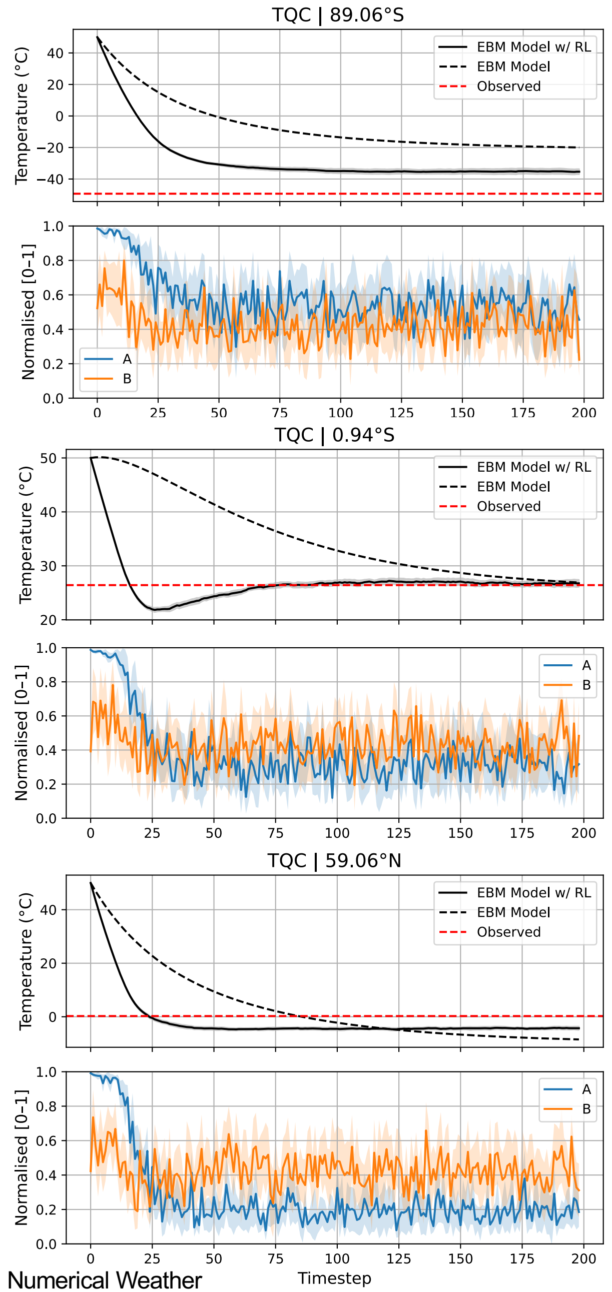
1. **DEMONSTRATED** faster convergence and geographically adaptive, state-dependent parameterisations through federated learning and spatial decomposition.
2. **DEVELOPED** in line with the modular design in GCMs, allowing region-specific corrections under global coordination via FedRAIN-Lite.
3. **ESTABLISHED** DDPG as a robust and efficient baseline, achieving stable convergence and low zonal errors across all setups.
4. **PROVIDE** a scalable prototype from idealised EBMs to operational GCMs, supporting data-driven and physically consistent climate modelling.











UNIVERSITY OF CAMBRIDGE

[1] Nath P, Schemm S, Moss H, Haynes P, Shuckburgh E, Webb M. FedRAIN-Lite: Federated Reinforcement Algorithms for Improving Idealised Numerical Weather and Climate Models. arXiv; 2025 NeurIPS workshop on Tackling Climate Change with Machine Learning. Available from: https://arxiv.org/abs/2508.14315.

[2] Nath P, Schemm S, Moss H, Haynes P, Shuckburgh E, Webb M. Making Tunable Parameters State-Dependent in Weather and Climate Models with Reinforcement Learning (Manuscript in preparation).

Next Steps …

MOVING TO THE

Unified Model

UM

Met Office